STRUCTURE-FUNCTION ANALYSIS AND GENETIC INTERACTIONS OF LUC7 AND MUD1 SUBUNITS OF SACCHAROMYCES CEREVISIAE U1 SMALL NUCLEAR RIBONUCLEOPROTEIN

A Dissertation
Presented to the Faculty of the Weill Cornell Graduate School
of Medical Sciences
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

By
Radhika Agarwal
December 2018

STRUCTURE-FUNCTION ANALYSIS AND GENETIC INTERACTIONS OF LUC7 AND MUD1 SUBUNITS OF SACCHAROMYCES CEREVISIAE U1 SMALL NUCLEAR RIBONUCLEOPROTEIN

Radhika Agarwal, Ph.D.

Cornell University, 2018

The discovery of the spliceosome, a large molecular complex composed of five small nuclear ribonucleoprotein particles (snRNPs) functioning in the removal of introns from precursor messenger RNAs stimulated an extensive field of research with innumerable questions on the mechanistic role of each component of the spliceosome.

U1 snRNP plays a crucial role in 5' splice site selection and initiates spliceosome assembly for pre-mRNA splicing in *S. cerevisiae.* The composition of U1 snRNP in *S. cerevisiae* is much more complex than in humans. This thesis focusses on establishing structure-function relations for Luc7, a 261 amino acid essential subunit of U1 snRNP, by conducting *in vivo* mutational analysis entailing N- and C-terminal truncations and alanine scanning of phylogenetically conserved amino acids, including two putative zinc-finger motifs, ZnF1 and ZnF2, and charged amino acids within ZnF2 module. I identified Luc7-(31-246) as a minimal functional protein. Moreover, mutations in ZnF2 module were found to be lethal which is now explained from structural insights to be due to the H-bonds formed by amino acids in this zinc-finger domain with U1 snRNA: pre-mRNA duplex. I also determined functional roles of otherwise benign mutations and deletion where they impaired splicing of a pre-mRNA with a nonconsensus splice site and were synthetically lethal with other splicing factors.

Mud1, an inessential 298-amino acid subunit of the *S cerevisiae* U1 snRNP. Mud1 consists of N-terminal and C-terminal RRM domains (RRM1 and RRM2) separated by a linker domain. This thesis dissects genetically the domain requirements of Mud1 by synthetic lethal interactions of *mud1*Δ with deletions or mutations in other spliceosome components. Interestingly, I found that all biological activities of Mud1 can be complemented by co-expressing separately the RRM1 (aa 1-127) and linker-RRM2 (aa 128-298) modules. I also investigated the probable role of the linker region and determined it to be a nuclear localizing signal.

Thus, the findings in this thesis highlight the contributions of Luc7 and Mud1 in stabilizing the interactions of U1 snRNP with the pre-mRNA 5' splice site.

BIOGRAPHICAL SKETCH

Radhika Agarwal received her Bachelor of Science degree in microbiology from St. Xavier's College, Kolkata University, India in 2008. She then received her Master of Science degree in biomedical genetics from VIT University, Vellore, India in 2010. She joined the BCMB Allied program at Weill Cornell Graduate School in August 2011 and has been a graduate student in the laboratory of Dr. Stewart Shuman at Sloan-Kettering Institute since September 2012.

To Mom and Dad

For teaching me to constantly strive to be better

To Amit

For always being there

ACKNOWLEDGMENTS

Seven years ago, when I left my home country India to embark on this adventurous journey I was anxious and excited at the same time. Anxious for leaving my family and friends back home and exhilarated to learn and experience the adventures of science while acquiring some life changing lessons on the way. I am thankful to all my mentors, colleagues, friends and family who made this journey worth remembering.

First, I am thankful to my mentor Stewart Shuman who has over the years helped me learn the art of thinking critically and seeking answers to my research questions. It is through his constant guidance, motivation and support that I have been able to achieve my goals. I am also grateful to my committee member Beate Schwer who has been more of a co-mentor and helped me understand the basics of yeast genetics and guided me at every step in my projects.

I am appreciative of Dirk Remus who has during committee meetings given some interesting insights that have helped me drive my research projects forward. I am also thankful to Kirk Deitsch, for his constant guidance during my graduate school career and for being the chair of my Thesis Defense.

I am thankful to my Master's molecular Biology teachers Dr. Pravir Kumar and Dr. Rashmi Ambastha who encouraged me to pursue a career in research and guided me along the way.

I am especially appreciative of Poulami Samai, a former graduate student in the lab, who was my mentor during my rotation in lab. She was extremely patient in teaching me the basics of protein purification, thinking through biochemical assays and performing them meticulously. Paul Smith, a former senior researcher in the lab was always there to help me troubleshoot through experiments and think of interesting ideas

during my ACE. I am appreciative of Anupam Chakravarty, a former graduate student in the lab whose mentorship as I started my thesis project in lab was immensely helpful. I am thankful to all the former lab members Heather Ordonez, Mathieu Chauleau, Lidia Chico, Agata Jacewicz, Ushati Das, William Maughan and Debashree Chatterjee. I am also appreciative of the current lab member Anam Ejaz, Angad Garg, Ankan Banerjee, Annum Munir, Barbara Remus, Brad Schmier, Loressa Uson, Mihaela Sandu, Shreya Ghosh, Weiyi Wang and Peihong Dai for their support and for providing a fun working environment in lab. I am happy to have made some great friends in my time in lab.

I am appreciative of Micheil Boekhout, a Research fellow at The Scott Keeney lab at Memorial Sloan Kettering who taught me the basics of microscopy with utmost clarity and guided me during some of my experiments.

I am thankful to my friends Pallavi Lamba, Shashirekha Mundhra and Bharat Vaidyanathan. We started our PhD together and their constant support made each hurdle easier to surmount. I am also grateful to my friends Nupur Khaitan and Dimple Hemani for their contribution in various ways. I am appreciative of my Uncle Narayan and cousin Premlata for their guidance during the start of my PhD.

Most importantly, my heartfelt gratitude goes to my family who has always been there to encourage, motivate and guide. My elder brother, Ram Jalan and sister in law, Rachna Jalan my niece, Ramana and my in-laws have been a constant source of support.

Lastly, I am indebted to my Mom, Suman Jalan and Dad, Ramesh Jalan who have always believed in me, showed me the right path and enabled me to take courageous steps in life. I am also thankful to have a lifelong friend in my husband Amit Agarwal, who has been a companion in true sense with his unceasing support and encouragement. This grad school experience comes with an added joy of raising my

daughter Amayra and I am grateful for all the support I have received from my mentor and family in this period.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

LIST OF ABBREVIATIONS

DNA: Deoxyribonucleic acid

RNA: Ribonucleic acid

mRNA: messenger RNA

hnRNA: heterogenous nuclear RNA

Ad2: Adenovirus2

ts: temperature sensitive

cs: cold sensitive

prp mutants: pre-mRNA processing mutants

BRR: Bad Response to Refrigeration

tRNA: Transfer RNA

nt: nucleotide

SS: Splice Site

BP: Branch Point

RNP: Ribonucleoprotein

snRNA: small nuclear RNA

Lsm: like Sm

TMG: 2,2,7-trimethylguanosine

snRNP: small nuclear ribonucleoprotein

CBC: Cap Binding Complex

NTC: Nineteen Compelx

ATP: Adenosine 5'-triphosphate

bp: base pair

NTP: Nuceloside 5'-triphosphate

2′-OH: 2′-hydroxyl

3′-OH: 3′-hydroxyl

Dbr1: Debranchase

LR: long range

SL: stem loop

MUD: Mutant U1 Die

LUC: Lethal Unless CBC

CBP: Cap binding protein

ZnF: Zinc-finger

RRM: RNA recognition motif

PCR: Polymerase chain reaction

ORF: Open reading frame

YPD: Yeast peptone dextrose

aa: amino acid

RT-PCR: Real time PCR

cDNA: complementary DNA

DMS: dimethylsulfate

GFP: Green fluorescent protein

SD: Synthetic Dropout

His: Histidine

PBS: Phosphate buffer saline

DAPI: 4',6-diamidino-2-phenylindole

NLS: Nuclear localization signal

5-FOA: 5-Fluoroorotic acid

PDB: RCSB protein data bank

5′-OH: 5′-hydroxyl

Chapter 1

INTRODUCTION

Discovery of Introns and Splicing

Until the 1960s genome architecture was believed to be a continuous stretch of DNA coding for proteins. This was true for prokaryotes, but it failed to explain the existence of large amounts of DNA in eukaryotic genomes. Moreover, researchers were in the quest to determine what resulted in the existence of long RNA (heterogeneous nuclear RNA; hnRNA) in the nucleus while the cytoplasm had shorter messenger RNA (mRNA) in vertebrates. Both of these RNA species were found to have a 7-methylguanosine cap structure at the 5' end and a polyadenylated 3' end. These observations lead to a model that hnRNA was a precursor of the mRNA produced via processing events during or after transcription (Darnell et al., 1971; Edmonds et al., 1971; Perry and Kelley, 1976; Rottman et al., 1974)

In 1977, Philip Sharp and Richard J. Roberts independently discovered the existence of introns and the process of splicing in a DNA virus, Adenovirus2(Ad2), that infects and replicates in mammalian cells. The adenovirus genome was known to be one long stretch of DNA. Electron microscopy mapping studies were used by Sharp and Roberts to compare the structure of the Ad2 DNA with mRNA. Adenovirus hexon, the major coat protein, was chosen for this study as it was known to be encoded by the most abundant late viral mRNA. It was discovered that RNA:DNA hybrids were seen in discontinuous segments with a region of the DNA looping out, called RNA displacement loops or R-loops. This led them to propose that the coding regions in the DNA (now called exons) are interspersed with non-coding sequences (now called introns). They further postulated that mature mRNA is produced by splicing of multiple

segments of a precursor nuclear RNA which also explained the existence of common 5' caps and 3' polyadenylated ends in hnRNA and mRNA(Berget et al., 1977; Chow et al., 1977). Similar attributes were soon observed in the other eukaryotic genes like those for ovalbumin and β-globin. Consensus sequences at exon-intron boundaries were also identified (Breathnach and Chambon, 1981). Roberts and Sharp later shared the Nobel Prize in 1993 for the discovery of "split genes".

The introduction to this thesis will provide an overview of the process of RNA splicing and the protein factors involved in the process with a focus on *Saccharomyces cerevisiae.* In particular, the first step of spliceosome assembly will be discussed in detail. Structure and function studies and genetic interactions of protein factors involved in spliceosome assembly will also be discussed. Comparisons will be drawn to the human/metazoan splicing machinery where relevant within the scope of this thesis.

Splicing in S. cerevisiae

Genetic tools to study splicing

The spliceosome, one of the most complex cellular machines involving the action of numerous proteins and RNAs that function in coordination to bring about precise cleavage of the pre-mRNAs, has been of immense interest to researchers since its discovery. Some questions that are particularly intriguing to us and others are: what is the role of each of the proteins in the complex and what is their mechanism of action? Functional conservation of the process between metazoans and single-celled eukaryotes has made *S. cerevisiae* an appropriate model system to study splicing. Genetic screens, e.g. temperature sensitive (ts) /cold sensitive (cs) enhancer and suppressor screens, have been a useful tool used for identification of genes and characterization of the proteins involved in splicing (Hossain and Johnson, 2014).

The foundation to study splicing genetically was laid by the temperature-sensitive mutant screens that identified a set of mutants that were defective in the formation of the ribosome and in splicing (Hartwell, 1967; Hartwell et al., 1970; Rosbash et al., 1981). These mutants are now known as the *prp (pre-mRNA processing)* mutants. Many of these PRP genes have been found to be essential for *in vitro* splicing and are known to play critical roles *in vivo* (Lustig et al., 1986). Subsequently, cold-sensitive screens identified 5 new genes BRR 1-5 (Bad Response to Refrigeration) that played important roles in splicing (Noble and Guthrie, 1996).

Enhancer and suppressor screens are indispensable to determine genetic interactions, which involve determining the effect of deletion in two non-essential genes or conditional mutations in essential genes. When double deletions/mutations are sicker than the single ones, it is referred to as synthetic growth defect or lethality (if death results from double deletions/mutations). The synthetic lethal phenotype highlights the importance of an interaction between two genes that is essential for viability. Also, the fact, that almost 80% of null alleles in *S. cerevisiae* are viable, points to the existence of buffering by the function of other genes, thus highlighting the concept of genetic redundancy (Boone et al., 2007; Hossain and Johnson, 2014). These genetic tools are the underpinnings of this thesis and have aided in the study of the U1 snRNP subunits, Luc7 and Mud1.

Characteristics of introns

Five classes of introns are identified in *S. cerevisiae* which are classified based on their excision mechanism, namely, (i) Group I introns that are spliced by two transesterification reactions initiated by a guanosine, (ii) Group II introns are also spliced by two transesterification reactions similar to that for nuclear pre-mRNAs. Some of the Group I and II introns are known to be autocatalytic and their splicing is

dependent on proteins, (iii) introns of the nuclear pre-mRNAs which are spliced by two transesterification reactions catalyzed by a multi-megadalton ribonucleoprotein complex called spliceosome, (iv) introns of the nuclear tRNA genes are removed by a heterotetrametric splicing endonuclease complex and (v) a distinct intron in the HAC1 mRNA whose splicing is catalyzed by Ire1p and tRNA ligase (Lopez and Seraphin, 2000).

Sequencing of fungal genomes has assisted in a thorough analysis of the number of introns, their position and how they evolved relative to multicellular eukaryotes. We now know that *S. cerevisiae* genes are intron poor with only 5% of them containing introns (Neuveglise et al., 2011), while approximately 95% of human genes contain introns (Venter et al., 2001). Out of the 5762 *S. cerevisiae* genes, only 287 are known to contain introns. Most of these genes contain only one intron and none of them have more than two introns. Interestingly, it has been observed that the introns in *S. cerevisiae* genes are concentrated at the 5' end (Neuveglise et al., 2011; Spingola et al., 1999). No such positional bias is observed in human genes which has an average of seven introns per gene distributed throughout (Neuveglise et al., 2011).

Another remarkable feature observed in *S. cerevisiae* is that ribosomal protein genes contribute to almost 90% of the mRNAs produced from intron-containing genes (Ares et al., 1999). This observation has been explained by Vinogradov (2001) that the introns in ribosomal protein coding genes have been retained due to selection pressure and they contribute to the regulation of ribosomal biogenesis.

Studies have shown that *S. cerevisiae* genes have a bimodal distribution of intron size. The longer introns i.e. 250-400 nt have been observed in genes encoding ribosomal proteins whereas the shorter introns i.e. 50-100 nt have been seen in non-ribosomal protein encoding genes (Bon et al., 2003; Neuveglise et al., 2011; Spingola et al., 1999).

Spliceosomal introns have three characteristic sequences, namely 5' splice site (5'SS), branch point (BP) and 3' splice site (3'SS). The consensus 5'SS for *S. cerevisiae* is GTATGT with exceptions seen in five introns that begin with GC. The branchpoint sequence, typically located 18-40 nt upstream of the 3'SS, in *S. cerevisiae* is defined as TACTAAC (A is the branchpoint nucleotide involved in the chemical steps of splicing as explained below) and the 3'SS sequence is defined as YAG (where Y is a pyrimidine) (Neuveglise et al., 2011; Spingola et al., 1999) (Fig. 1.1).

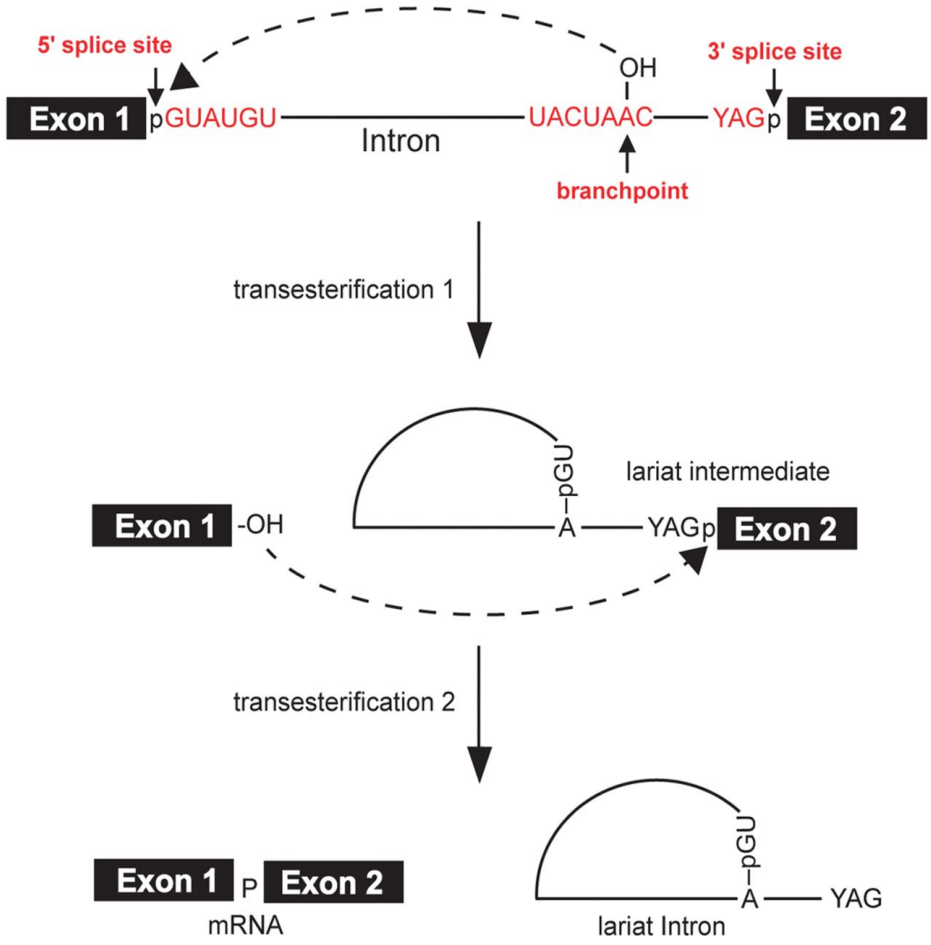


Figure 1.1. Catalytic steps in splicing. *Saccharomyces cerevisiae* conserved splice site sequences are shown in red. Splicing involves two transesterification reactions. In the first step 5' splice site is cleaved and a lariat intron-exon2 intermediate with a 2'-5' phosphodiester bond is formed. The second step involves ligation of Exon1 and Exon2 and release of the intron lariat (Chen and Cheng 2012).

Composition of The Spliceosome

Eukaryotic pre-mRNA splicing is catalyzed by a multi-megadalton ribonucleoprotein (RNP) complex called the spliceosome. The spliceosome is composed of 5 uridine-rich small nuclear RNAs (snRNAs) namely, U1, U2, U4, U5 and U6. U1-U5 snRNAs are associated with a 7-subunit Sm-protein ring (B, D1, D2, D3, E, F, G). U6 snRNA binds a homologous 7-subunit Lsm 2-8 (like Sm) protein ring. All of the snRNAs except U6 have a 2,2,7-trimethylguanosine (TMG) cap at the 5' end (Guthrie and Patterson, 1988). Also, each of the snRNAs associates with specific proteins to form small nuclear ribonucleoproteins (snRNPs) (Table 1.1) (Luhrmann et al., 1990; Raker et al., 1996). U4 base pairs with U6 to form U4/U6 di-snRNP followed by interaction with U5, leading to the formation of U4/U6•U5 tri-snRNP (Bringmann et al., 1984; Konarska and Sharp, 1987). In addition, a larger number of proteins which include over 30 *PRP* genes have been identified that are involved in the process of splicing but are not directly associated with the snRNPs. DExD/H-box RNA helicases, are one of such class of proteins (Table 1.2) (Cordin and Beggs, 2013). Other non-snRNP proteins include: Msl5 and Mud2, which are involved in binding pre-mRNA at the branchpoint, CBC1 and CBC2, which form a nuclear cap binding complex, and the yeast nineteen complex (NTC). Thus, the spliceosome is the largest cellular machine formed by an intricate network of RNA-RNA, RNA-protein, and protein-protein interactions (Nilsen, 2003). It is noteworthy that a complete splicing cycle requires 8 different helicases, thus highlighting the dynamic nature and need for conformational changes in the snRNAs during pre-mRNA splicing (Madhani and Guthrie, 1994).

Table 1.1. Protein subunits of the *S. cerevisiae* snRNPs. (Brow 2002)

U1	U2	U4	U5	U6	U4/U6	U4/U5/U6
Sm B-G	Sm B-G	Sm B-G	Sm B-G	Lsm 2-8	Prp6	Prp38
Snp1	Lea1	Prp3	Dib1	Prp24	Prp31	Snu23
Mud1	Msl1	Prp4	Prp8			Snu66
Yhc1	Cus1	Snu13	Prp44			Spp381
Luc7	Cus2		Snu40			
Nam8	Hsh49		Snu114			
Prp39	Hsh155					
Prp40	Prp9					
Prp42	Prp11					
Snu56	Prp21					
Snu71	Rse1					
	Snu17					

Table 1.2. Spliceosomal DExD/H-box RNA helicases. (Adapted from Brow 2002)

DExD/H-box protein	Human homolog	Stage in splicing
Sub2	UAP56	Pre-spliceosome
Prp5	KIAA0801	Pre-spliceosome
Prp28	U5-100kD	Early step I activation
Prp44/Brr2	U5-200kD	Early step I activation
Prp2	KIAA0577	Late step I activation
Prp16	hPrp16	Step II activation
Prp22	Hrh1	mRNA release
Prp43	DDX15/hPrp43	Intron release

Steps involved in the splicing-reaction cycle

The process of pre-mRNA splicing is broadly divided into 4 steps: (i) assembly of the core splicing machinery, (ii) spliceosome activation, (iii) catalysis which involves two transesterification reactions and spliceosome disassembly. The major snRNAs and protein components involved in this process are described below based on the *S. cerevisiae* nomenclature.

Assembly

Spliceosome assembly is initiated by binding of U1 snRNP to 5'SS via base pairing interaction between the conserved hexanucleotide motif 5'-ACUUAC on the U1 snRNA and the consensus 5'SS element 5'-GUAUGU. This is an ATP independent process and leads to the formation of an early assembly intermediate (Complex E) (Fig. 1.2). The Msl5-Mud2 heterodimer then binds to the branch point sequence before any of the other splicing complexes interacts with the pre-mRNA. This early spliceosome is stabilized by bridging interactions between U1snRNP and Msl5-Mud2 (Abovich and Rosbash, 1997) and it ensures the commitment of the pre-mRNA to the splicing pathway. The first ATP-dependent step in the splicing pathway involves association of the U2 snRNP (Crawford et al., 2013). Two DExD/H helicases (Table 1.2) Sub2 and Prp5 are involved in this process (Fig. 1.2). Generation of a catalytically active spliceosome entails displacement of Msl5-Mud2 from the branchpoint such that U2 snRNA can base pair with the intron branchpoint UACUAAC. Sub2 is postulated to play a role in this process, disrupting or remodeling contacts mediated by Mud2 while simultaneously assisting in the binding of U2 snRNP at the branch site forming the pre-spliceosome (Complex A) (Kistler and Guthrie, 2001; Libri et al., 2001; Zhang and Green, 2001).

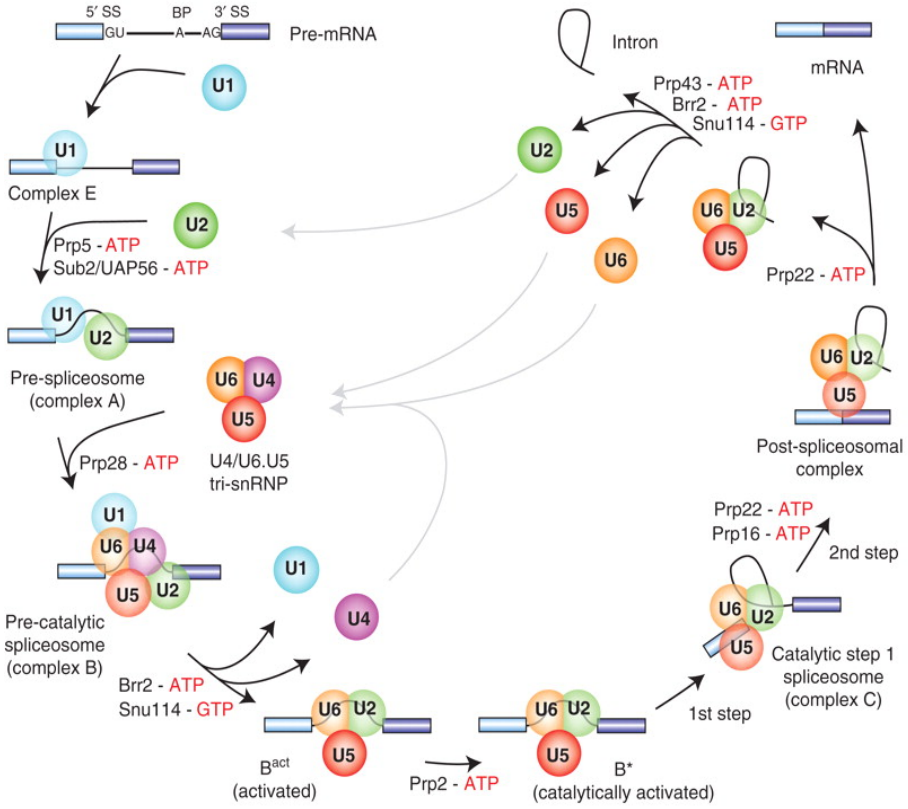


Figure 1.2. Schematic representation of the *S. cerevisiae* spliceosome pathway. There are four stages in the pathway: assembly, activation, catalysis and disassembly. Interactions of snRNPs with the pre-mRNA leads to spliceosome assembly. Release of U1 and U4 snRNPs along with binding of NTC triggers spliceosome activation which is followed by the two catalytic steps of the splicing reaction. In the last stage the mature mRNA is released, intron is debranched and targeted for degradation, and the spliceosome components are disassembled (Will and Luhrmann, 2011)

Prp5 is proposed to be responsible for conformational changes in U2 snRNA resulting in an increased accessibility of its branchpoint binding sequence (O'Day et al., 1996).

U4 snRNP is typically known to exist in association with U6 via extensive base pairing interactions involving 24 base pairs (bp) forming U4/U6 di-snRNP or U4/U6•U5 tri-snRNP (Brow and Guthrie, 1988; Cheng and Abelson, 1987). Formation of the free U5 snRNP involves the role of three proteins: Prp8, Brr2 and Snu114. Prp8 interacts with U5 snRNA and is stabilized by interaction with GTP bound Snu114 (Fabrizio et al., 1997). This is followed by association of Brr2 with the complex leading to a fully formed U5 snRNP. U4/U6 di-snRNP then associates with this free U5 snRNP particle forming the U4/U6•U5 tri-snRNP that binds to the spliceosome leading to the formation of pre-catalytic B complex (Cheng and Abelson, 1987; Stevens et al., 2001). (Fig 1.2)

Activation

The foremost rearrangement and conformational change that occurs during spliceosome activation involves replacement of U1 with U6 at the 5'SS (Staley and Guthrie, 1999). This process involves the role of an essential DExD/H RNA helicase Prp28, which is discussed below. Also, disruption of the U4/U6 di-snRNP initiates base pairing between U6 and U2 snRNAs. An extensive network of protein interactions involving Prp8, Brr2 and Snu114 is involved in regulating this process (Bartels et al., 2003; Nielsen and Staley, 2012; Small et al., 2006). After the release of U1 and U4, the Nineteen complex (NTC) composed of Prp19 and seven associated proteins binds to the spliceosome stabilizing the interactions between U6 and U5 (Chan and Cheng, 2005; Chan et al., 2003). The catalytic core of the spliceosome is formed by base pairs between U2 and U6 snRNA. U4 snRNA while remaining in association with U6 snRNA ensures that this catalytically activated spliceosome B^*complex is not formed prematurely (Brow and Guthrie, 1989) (Fig 1.2).

Prp28 – role in activation of spliceosome

Strauss and Guthrie in 1991 were in the quest to identify new trans-acting proteins involved in pre-mRNA splicing. Assembly of macromolecules was considered to be especially sensitive to low temperature due to the fact that most of the interactions driving this process are hydrophobic in nature and low temperatures mutations affecting this assembly process would show a more pronounced phenotype. Also, certain genes are known to mutate preferentially to cold-sensitive alleles thus helping in the identification of genes that may otherwise be missed. Strauss and Guthrie isolated Prp28 as a cold-sensitive splicing mutant (Strauss and Guthrie, 1991). It was defined to belong to the DEAD-box family of RNA helicases. Members of this family are identified by the presence of 12 conserved amino-acid sequence motifs. This class of RNA helicases in addition to playing a role in RNA-binding and unwinding and NTP hydrolysis, are now believed to have more diverse roles. For example, some of these DEAD-box family of proteins have been determined to function more locally in affecting the RNA conformation whereby they weaken or strengthen RNA-protein or RNA-RNA interactions. (Linder and Fuller-Pace, 2013; Putnam and Jankowsky, 2013).

Prp28 in *S. cerevisiae* is essential and known to have an RNA-dependent ATPase activity which is rather feeble with respect to ATP turnover number (Jacewicz et al., 2014). Certain DEAD-box proteins are known to function as an RNA switch rather than as translocases where ATP binding helps in creating a RNA or protein distortion and ATP hydrolysis plays a role in recycling the protein (Del Campo and Lambowitz, 2009; Linder and Fuller-Pace, 2013; Putnam and Jankowsky, 2013). Similarly, it is postulated that the weak ATPase activity of Prp28 makes it suitable to function as an RNA switch allowing replacement of U1 snRNP with U6 snRNP at the 5'SS (Jacewicz et al., 2014).

Also, genetic studies have shown that mutations in essential U1 snRNP subunits Yhc1, Prp42, Snu71 and SmD2 and specific mutations in the U1 snRNA within regions that base pair with 5'SS, result in the bypass of the requirement of Prp28. I have shown that Luc7 mutations (discussed in Chapter 2 and 4) bypass the Prp28 requirement indicating that the need for Prp28 is assuaged when U1•5'SS contacts are weakened due to mutations in factors responsible for stabilizing it (Chen et al., 2001; Hage et al., 2009; Schwer et al., 2013; Schwer and Shuman, 2014; Staley and Guthrie, 1999). Therefore, these findings support the hypothesis that Prp28 plays a role in the conformational switch during spliceosome assembly.

Catalytic steps in splicing

Pre-mRNA splicing is carried out by two S_N2-type transesterification reactions. In the first step, the 2'OH of the branch-point adenosine carries out a nucleophilic attack on the phosphate of the guanine nucleotide at the 5'SS. This results in the cleavage of a 3'-5' phosphodiester bond. Simultaneously, the first nucleotide of the intron and branch point adenosine forms an unusual 2'-5' phosphodiester bond. In the second step, the 3'-OH of the 5' exon (released by the first transesterification reaction) carries out a nucleophilic attack on the phosphate at the 3'SS. This results in the ligation of the exons through a 3'-5' phosphodiester bond. The intron is released in the form a lariat – a lasso like intron structure (Chen and Cheng, 2012; Shi, 2017) (Fig. 1.1).

Group II self-splicing introns also produce similar intermediates and products as described above for pre-mRNA splicing in *S. cerevisiae.* Therefore, it is postulated that pre-mRNA splicing is catalyzed by RNAs similar to group II self-splicing introns (Gordon et al., 2000). It is proposed that they both follow a Mg^{2+} dependent general two-metal catalytic mechanism where the snRNA coordinates the catalytic metal ions (Steitz and Steitz, 1993). One metal ion is implicated to activate the sugar hydroxyl

while the other metal ion directly coordinates and stabilizes the oxyanion leaving group. In pre-mRNA splicing these metal ions are believed to be coordinated by the phosphates in U6 snRNA. Three metal binding sites are known to be present in U6 snRNA, namely an AGC triad, position U80 in the internal loop and the third at ACAGAGA sequence which is known to base pair with the 5'SS (Fabrizio and Abelson, 1992; Lee et al., 2010). Structural constraints ensure that these metal binding sites are in close proximity to the reactive groups for the first chemical step of splicing. A likely spliceosomal remodeling then occurs before the second transesterification reaction (Rhode et al., 2006). Although the exact role of Mg^{2+} ions in the catalytic steps is not known, it has been shown that sulphur substitution of the non-bridging phosphoryl oxygen at position U80 in U6 snRNA resulted in a fully assembled but catalytically inactive spliceosome. This defect was rescued by using a more thiophilic metal like Mn^{2+} (Yean et al., 2000).

Prp2, Brr2 and Prp16 are the ATPases that play a significant role in catalytic steps of splicing along with other proteins that are part of the NTC (Fig. 1.2). Before the first catalytic step Prp2 interacts with the intron and the carboxyl terminal of Brr2 (Liu and Cheng, 2012), this activates the ATPase activity of Prp2 leading to the displacement of a majority of U2 snRNP subunits thereby unmasking the 2' hydroxyl at the branchpoint. This places the 2' hydroxyl in a favorable conformation for an in-line nucleophilic attack on the phosphate of the 5'SS, which is ATP-independent (Lardelli et al., 2010).

The second catalytic step includes the recognition of the 3'SS for its alignment with the 5'SS leading to exon ligation. Conserved nucleotides in U5 snRNA have been shown to interact with the exons thus facilitating their alignment (Newman and Norman, 1992). It is postulated that the U5 interaction with the exons is further stabilized by U5 snRNP subunit Prp8 based on the findings that Prp8 cross-links with the two exons

proximal to the splice-site junction (Dix et al., 1998; Teigelkamp et al., 1995). DEAH-box RNA helicase Prp16 is hinted at playing an important role in remodeling of the spliceosome after the first catalytic step (Schwer and Guthrie, 1992). This role of Prp16 was suggested based on the observation that Prp16 action resulted in RNase-H protection of the 3'SS (Schwer and Guthrie, 1992) which was later proposed to be due to transient association of Prp22, a DEAH-box ATPase/helicase at the 3'SS (McPheeters et al., 2000) (Fig. 1.2). Suppression of mutations in RNA duplex forming regions in U2 and U6 snRNAs in the presence of cold-sensitive *prp16* alleles (Mefford and Staley, 2009) suggests that Prp16 plays a role in remodeling the snRNA interactions, specifically by disrupting RNA duplex interactions between U2 and U6 snRNA. Also, Prp16 was indicated to be involved in an ATP-dependent release of first step splicing factors Cwc25 and Yju2 thus setting the stage for the second transesterification reaction to occur with the help of other associated factors as discussed below (Tseng et al., 2011).

Subsequent to Prp16 action, Slu7, Prp18 and Prp22 are required for the second transesterification reaction when the distance between the branchpoint and the 3'SS is more than seven nucleotides (Brys and Schwer, 1996; Horowitz and Abelson, 1993; Ohrt et al., 2013; Schwer and Gross, 1998). The three proteins aid in positioning the 3'SS and the 3'OH of the 5'exon in the active site (Schwer, 2008). Prp22 acts in the second catalytic step in an ATP-independent manner (Schwer and Gross, 1998). After the completion of the splicing reaction, Prp22 unwinds the mRNA/U5 snRNA duplex and releases the mature mRNA via ATP hydrolysis (Schwer, 2008; Schwer and Gross, 1998) (Fig. 1.2).

Disassembly

Spliceosome disassembly involves release of the lariat-intron and all associated components for recycling. The lariat-intron is debranched by the enzyme debranchase

(Dbr1) which results in linearization of the intron that is eventually degraded (Chapman and Boeke, 1991). Prp43, an RNA helicase of the DEAH-box family, along with co-factors Ntr1 and Ntr2, forms an NTR complex involved in mediating the disassembly of the spliceosome components in an ATP-dependent process (Boon et al., 2006; Tsai et al., 2005) (Fig. 1.2). Mechanistic insights have been provided on the disassembly process where recruitment of Prp43 to the spliceosome takes place via interaction between Ntr2 and U5 snRNP subunit Brr2. However, Ntr2 binding to the spliceosome is inhibited in the presence of Prp16 and Slu7 which protects against premature spliceosome disassembly (Chen et al., 2013).

U1 snRNP

U1 snRNP in humans consists of the 2,2,7-trimethylguanosine (TMG)-capped U1 snRNA (164 nt), a Sm-protein ring with seven subunits and three U1-specific subunits U1-70K, U1-C and U1-A. U1 snRNP in yeast is much more complex in composition, consisting of a larger TMG-capped U1 snRNA (568 nt), conserved Sm-protein ring and homologs of the human U1-specific subunits, U1-A (Mud1), U1-C (Yhc1) and U1-70K (Snp1) along with 7 additional proteins namely, Prp39, Prp40, Snu71, Snu56, Luc7, Prp42 and Nam8. Amongst these proteins Nam8 and Mud1 are inessential and others are essential for viability.

U1 snRNA secondary structure

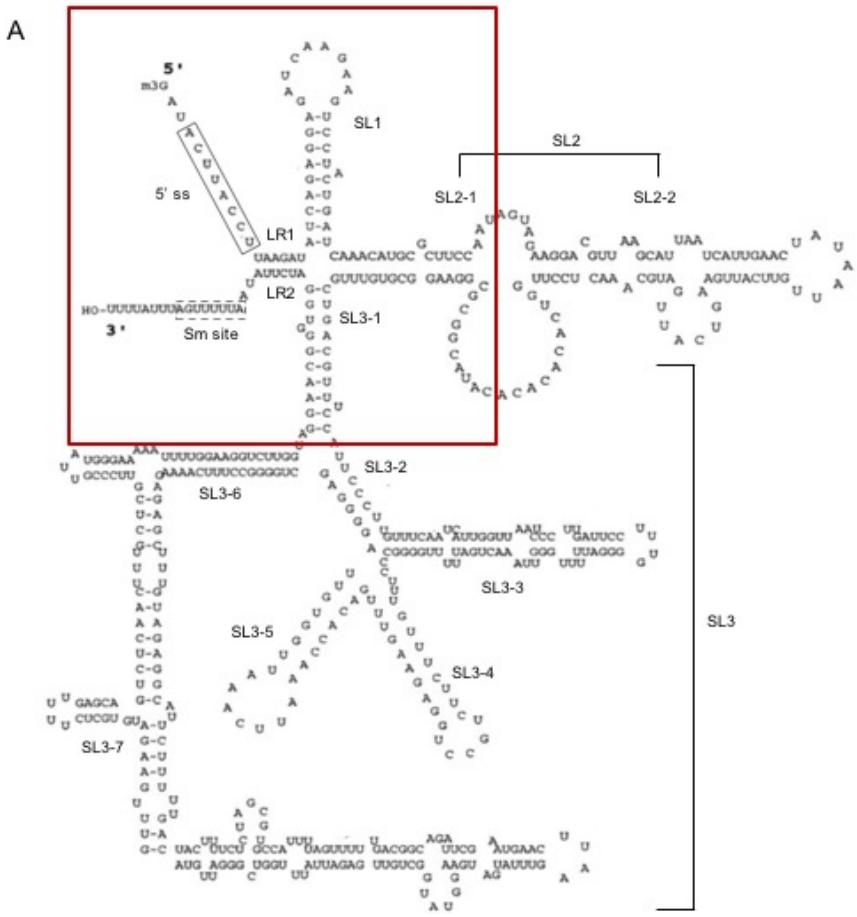
S. cerevisiae U1 snRNA with 568 nt is much larger than human U1 snRNA of 164 nt. Secondary structure predictions and recent cryo-EM data show that the *S. cerevisiae* U1 snRNA can be divided into the following elements: 5'ss recognition sequence, long-range interaction 1 (LR1), Stem-loop (SL) 1, SL2, SL3 (which is subdivided into SL3-1 to SL3-7), LR2, Sm-binding site and the 3'tail. U1 snRNAs from

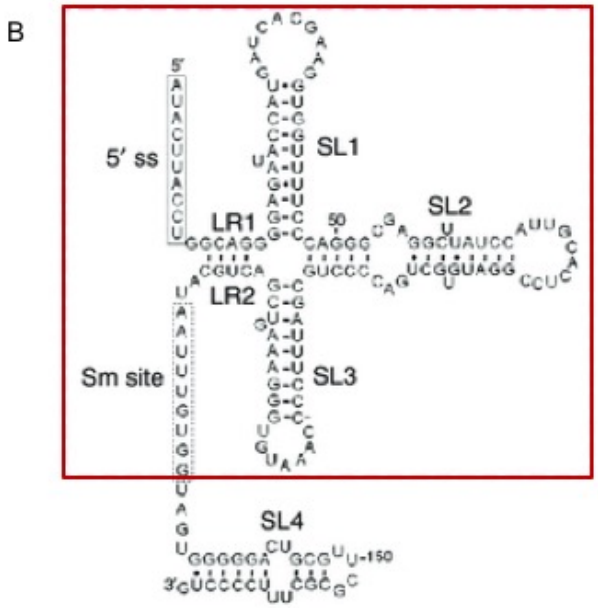
S. cerevisiae and humans have similar features in LR1, LR2, SL1 and Sm-binding site regions whereas the SL2, SL3 and 3'tail regions have significant differences (Kretzner et al., 1990; Li et al., 2017) (Fig. 1.3).

Genetic studies on U1 essentiality in *S. cerevisiae* demonstrated that a large portion of the conserved SL1 (also known as A loop) and *S. cerevisiae* specific SL3 region (also known as the core region) could be deleted with no defects in growth (Liao et al., 1990). However, U1 snRNA levels were found to be reduced 10-fold in these deletion strains thus suggesting that reduction of U1 snRNA levels up to 10-fold does not result in a corresponding defect on growth. To further test the effect of these deletions a reporter assay was used where an inefficiently spliced intron was inserted in the the β-galactosidase gene. The deletion strains of U1 snRNA results in reduced β-galactosidase expression thus indicating that these strains were defective in splicing an inefficiently spliced pre-mRNA. Moreover, growth rate defects were exacerbated when mutation combinations of SL1 and SL3 were tested suggesting that one mutation makes the cells sensitive to the presence of other mutations thus enhancing the growth phenotypes (Liao et al., 1990). Also, other than the conserved 5' and 3' ends deletion of upto 316 nt did not result in any growth defects, while deletion of 355 nt resulted in slow growth (Siliciano et al., 1991).

Figure 1.3. Comparison of secondary structure of *S. cerevisiae* and human U1 snRNA. (A) Secondary structure of *S. cerevisiae* U1 snRNA (Adapted from (Tang et al., 1997). (B) Secondary structure of human U1 snRNA. Red boxes indicate the common core structure between *S. cerevisiae* and human (Taken from (Li et al., 2017)

.


A
5'
m3G
5' ss
SL1
SL2
SL2-1
SL2-2
LR1
LR2
SL3-1
3'
Sm site
SL3-2
SL3-6
SL3-3
SL3-5
SL3-4
SL3-7
SL3


B
5' ss
SL1
SL2
LR1
LR2
Sm site
SL3
SL4

Identification of S. cerevisiae U1 snRNP subunits

Sequence analysis and homology studies led to the identification of the homologs of the human U1 snRNP subunits U1-70K and U1-C: Snp1 and Yhc1 respectively (Smith and Barrell, 1991; Tang et al., 1997). Sm D1, D3, E, F and G subunits of the heptameric Sm-rind were also identified via homology studies. The Sm B/B' subunit of the Sm-ring was identified through mass-spectrometry (Gottschalk et al., 1998).

Results on the inessentiality of large chunks of U1 snRNA, where null strains are inviable, enthused the development of a genetic screen for mutations that resulted in synthetic lethality with an otherwise benign mutation in U1 sRNA. This involved deletion of the large *S. cerevisiae* specific SL-3 region along with a point mutant G27A in SL-1 which was known to interact with U1-70K in human. Thus, this screen was named the MUD screen – Mutant U1 Die. This screen led to the identification of four U1 specific subunits: Mud1, Snu56/Mud10, Nam8/Mud15, Prp42/Mud16. A subunit of the nuclear cap binding complex Cbc2/Mud13 was also identified in this screen (Abovich et al., 1994; Colot et al., 1996; Gottschalk et al., 1998; Liao et al., 1993).

Prp40, an essential subunit of U1 snRNP, was identified in a suppressor screen where a cold sensitive mutation in the 5' end of U1 snRNA was made which resulted in death at 18˚C. *PRP40* gene was identified as a potential candidate such that it restored growth at 18˚C after mutagenesis for the strain carrying the U1snRNA mutation at its 5' end (Kao and Siliciano, 1996). Prp39, another essential U1 snRNP subunit was identified in a *ts* genetic screen for conditional mutants that were defective in splicing (Lockhart and Rymond, 1994). Both of these U1 snRNP subunits were shown to be essential for the interaction of the U1 snRNP with the 5'SS during spliceosome assembly (Kao and Siliciano, 1996; Lockhart and Rymond, 1994).

Cryo-EM structure of S. cerevisiae U1snRNP

The first Cryo-EM structure of *S. cerevisiae* U1 snRNP was determined by Li *et al* in 2017. They purified the yeast U1 snRNP with a TAP-tag on Mud1. The purified U1 snRNP complex was determined to contain all its 17 subunits and the U1 snRNA along with Sto1 and Cbc2. The cryo-EM structure was resolved to a resolution of 3.6 Å. The core/central region of the structure consisted of U1 snRNA, Sm-ring, Mud1, Yhc1 and Snp1 at similar positions as in the human cryo-EM structure of U1 snRNP (Li et al., 2017; Pomeranz Krummel et al., 2009). The core region is further seen to bind to the peripheral *S. cerevisiae* specific proteins Luc7, Nam8, Prp39, Prp42, Snu56, and Snu71. Two other cryo-EM structures of Pre-spliceosome A-complex inclusive of U1 snRNP and U2 snRNP along with pre-mRNA and Pre-B complex comprising of the A-complex and U4•U5•U6 tri-snRNP were published recently with similar U1 snRNP protein arrangement while also showing extensive contacts with pre-mRNA along with the U1 snRNA, thus giving mechanistic insights into 5'SS recognition (Bai et al., 2018; Plaschka et al., 2018) (Fig. 1.4).

The U1 snRNA SL-2 region was observed to interact with Mud1 RRM1. Mud1 was seen to bind CACAUAC sequence different from the AUUGCA sequence for U1A in humans. Extensive interactions are seen in Mud1 RRM1 with two internal loops in the SL2-1 and SL2-2 regions of U1 snRNA suggesting that Mud1 may be stabilizing the U1 snRNA. The contacts of Mud1 RRM1 with U1 snRNA are discussed in detail in Chapter 3. Yhc1with its extended C-terminal domain unlike the human U1-C makes extensive contacts with SmD3, Nam8 and Prp42. Snp1 was seen to have stabilizing interactions with Yhc1 via its long N-terminal arm which was wrapped around the Sm ring (Li et al., 2017).

The pre-mRNA 5'SS GUAUGU was observed to base pair with the 5' end of U1 snRNA ^{1}AUACUUACCUU^{11}. With this duplex formed, U1 snRNA and pre-mRNA are recognized by extensive atomic interactions with Yhc1, Luc7, SmB and SmD3 subunits of the U1 snRNP. The C-terminal Zn finger of Luc7, N-terminal Zn-finger domain of Yhc1, C-terminal residues of SmB and SmD form H-bonds to the phosphodiester backbone of U1 snRNA and pre-mRNA and specific H-bonds to some of the bases of the RNAs. Overall the positive surface potential of the surface amino acids of these proteins contributes to the charge neutralization and stabilization of the U1 snRNA and pre-mRNA duplex (Bai et al., 2018; Plaschka et al., 2018). Nam8 was seen to fit in a compact position between Prp42 and Prp39. Snu56 was observed to wrap around Prp42 between its N- and C-terminal domains. The common theme that emerged from the structures of U1 snRNP was that Prp42 along with its paralog Prp39 formed a central scaffold for the interaction and stabilization of the other U1 snRNP subunits (Bai et al., 2018; Li et al., 2017; Plaschka et al., 2018). Prp40 has not been modelled in any of the structures published.

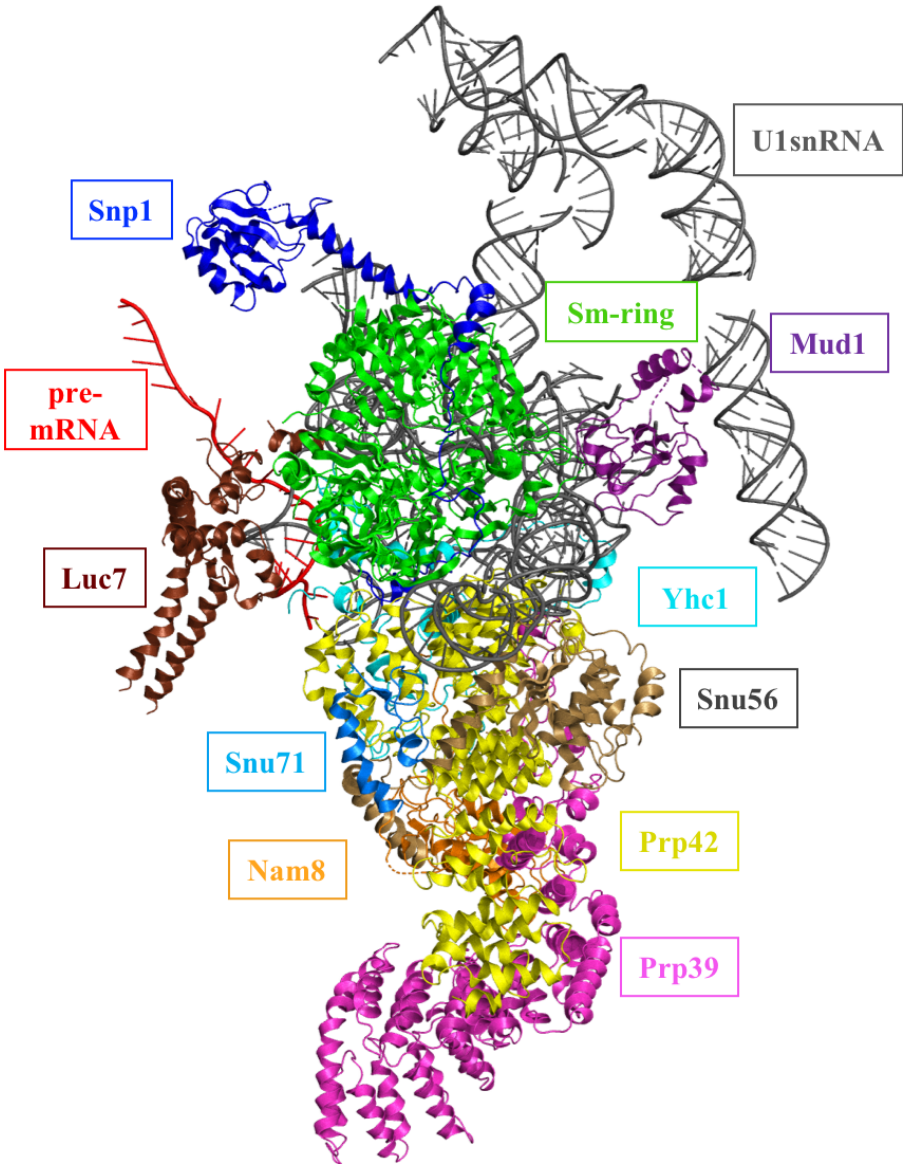


Figure 1.4. Structure of the *S. cerevisiae* U1 snRNP: Model of U1 snRNP was obtained from the Cryo-EM structure of the *S. cerevisiae* pre-B complex (PDB ID: 5ZWN). The U1 snRNA and pre-mRNA are colored grey and red respectively. The U1 snRNP subunits are each represented by a different color (Adapted from (Bai et al., 2018).

Trimethylguanosine synthase (Tgs1) and its role in synthetic interactions with U1 snRNP subunits

Newly formed RNA polymerase II transcripts are characterized by the presence of a 5' m^7GpppN (m^7G) cap structure via a three-step process catalyzed by three enzymes in *S. cerevisiae*: (i) RNA triphosphatase (Cet1) hydrolyzes the 5' triphosphate end of the RNA to a diphosphate, (ii) RNA guanylyltransferase (Ceg1) transfers a GMP from GTP to the diphosphate end of the RNA to form a 5' GpppRNA, and (iii) RNA methyltransferase (Abd1) methylates the cap guanine-N^7 position to form a m^7G cap. S-adenosyl methionine (AdoMet) acts as a methyl donor in the last step (Shuman, 2001). In addition to providing protection from 5' to 3' exonucleases, the m^7G caps are also known to trigger pre-mRNA splicing and RNA 3' end processing events. In vertebrates it also plays a role in nuclear export of the U snRNAs. However, it is not known if a parallel process exists in *S. cerevisiae* (Fortes et al., 1999) (Fig. 1.5).

In case of snRNAs the m^7G cap is further methylated to form 2,2,7-trimethylguanosine (TMG) caps by the enzyme Tgs1. Post transcription, Tgs1 performs two consecutive methyl transfer reactions from AdoMet to the N2 position of cap guanine (Fig. 1.5). The presence of a m7G cap on the RNA is essential for Tgs1 activity (Mouaikel et al., 2002).

A two-hybrid screen in *S. cerevisiae* for proteins interacting with the Sm-B subunit of the heptameric Sm-ring led to the identification of Tgs1. Tgs1 is non-essential in *S. cerevisiae* where a null strain results in cold sensitivity with a growth defect below 25˚C. Inessentiality of Tgs1 was surprising in as much as; though m7G caps are essential; TMG cap structures found to be conserved in eukaryotic small RNAs are not essential (Mouaikel et al., 2002). Thus, it was postulated that in the absence of Tgs1 there may be alternative pathways to ensure the function of TMG-capped RNAs in the

absence of Tgs1. Synthetic enhancement genetics were pursued to screen for mutations that led to lethality or heightened growth defects in *tgs1*Δ background thus reiterating the existence of cellular buffering under physiological conditions. Synthetic genetic interactions were observed with Tgs1 for U1 snRNP subunits Mud1 and Nam8, branchpoint binding protein Mud2, and U2 snRNP subunit Lea1 (Hausmann et al., 2008). These studies inspired me to test *tgs1*Δ strain for mutational synergies with the essential U1 snRNP subunit Luc7 as discussed in Chapter2.

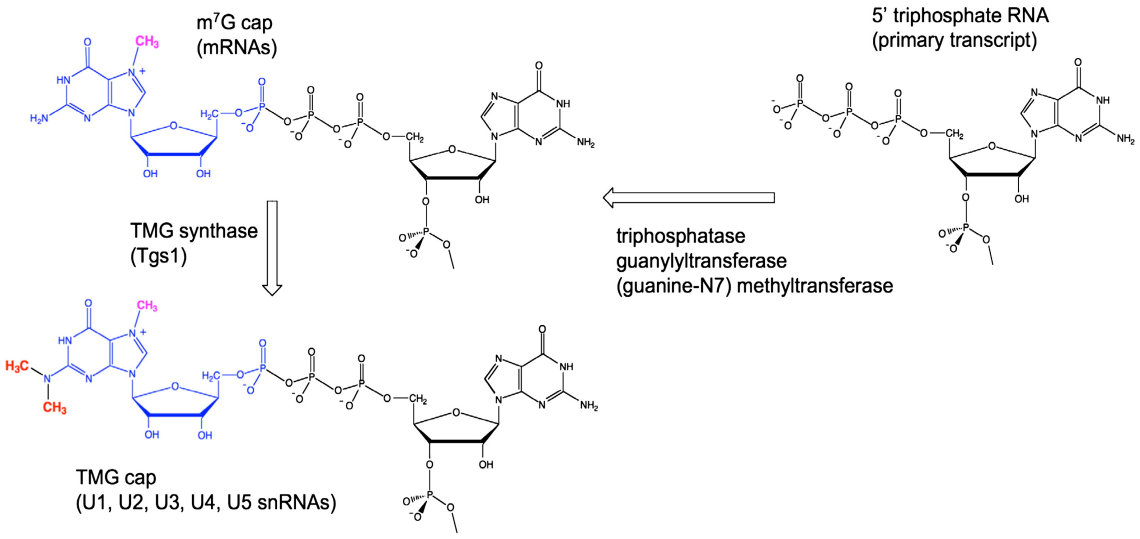


Figure 1.5. Synthesis of m^{7}G and TMG caps. 5' m^{7}G cap structure is formed by enzymes that act on the 5'triphosphate end of the RNA. The process involves removal of the γ-phosphate, addition of a guanylate (blue) to the β-phosphate and transfer of a methyl group (magenta) from AdoMet to the cap guanine-N7 position. The enzyme Tgs1 catalyzes the addition of two additional methyl groups (red) to the N2 position of cap guanine. (Adapted from Shuman, 2007)

Cap Binding Complex (CBC) and its role in interaction of U1 snRNA and 5'SS

Two cap binding protein complexes eIF4e and CBC facilitate the function of the m^7G caps. The cytoplasmic translation initiation factor eIF4E regulates the interaction of the 40S ribosomal subunit with 5' mRNA end. In *S. cerevisiae* eIF4E is encoded by the essential gene *CDC33* (Fortes et al., 1999). CBC is the nuclear cap binding complex that was identified to play a role in pre-mRNA splicing and nuclear export in metazoans (Izaurralde et al., 1995; Izaurralde et al., 1994). Studies indicated that CBC was required for the recognition of the 5'SS by the U1 snRNP (Lewis et al., 1996). Metazoan CBC is composed of two subunits: CBP80 and CBP20.

S. cerevisiae CBC subunits Sto1 (homologue of CBP80) encoded by the *GCR3* gene was identified as a submission of a known gene in GenEMBL by multiple groups and found to have 33% identity to human CBP80 (Izaurralde et al., 1994). Cbc2 (homologue of CBP20) was identified to be encoded by the *MUD13* gene in a synthetic screen for mutations that led to lethality in the presence of specific U1 snRNA mutations as described above. Pre-mRNA splicing defects were observed for *gcr3*Δ and *mud13*Δ strain (Colot et al., 1996). Further, genetic studies to define more precisely the role of CBC were undertaken. A Synthetic lethal screen identified the "LUC (Lethal Unless CBC) mutants" which resulted in death in the absence of CBC (Fortes et al., 1999). A great majority of these genes were found to be involved in spliceosome assembly and hence it is further suggestive of the role of CBC in modulating the interaction of U1 snRNP with the 5'SS (Fortes et al., 1999; Lewis et al., 1996).

The crystal structure of human CBP20/CBP80 highlighted some key features that explained the function of CBC (Mazza et al., 2002). CBP20 was seen to interact physically with the m^7G via the cap-binding pocket, CBP80 on the other hand had no direct contact with the cap but was detected to interact with the N-terminus of the

CBP20 thus stabilizing the cap-binding pocket. An important feature of the cap binding was the π-stacking interaction between the positively charged m^7G base and Tyr24 and Tyr49 of *S. cerevisiae* Cbc2. Alanine mutations of these tyrosines resulted in reduced affinity for m^7G cap binding (Mazza et al., 2002). Genetic interactions of the hypomorphic allele *Cbc2-Y24A* with pre-mRNA splicing factors identified subunits that when deleted or mutated elicited lethal or sick phenotypes thus highlighting cap-dependent functions of CBC (Qiu et al., 2012). I further exploited the knowledge from these genetic interactions to study if there were any genetically buffered functions between CBC and Luc7 as discussed in Chapter 2.

It is intriguing that while *S. cerevisiae* U1 snRNP is more complex in composition than human U1 snRNP, some of its protein subunits are inessential. This has encouraged a number of genetic studies to determine the existence of redundant factors in the process of splicing assembly. Redundancy here implies that there may be additional factors that can function in the absence of the existing factors and hence even though some of these components are inessential their functional role is magnified in synthetic backgrounds. An intricate network of genetically buffered functions have hence been identified that include U1 snRNP subunits Nam8, Yhc1, Snp1 and U1 snRNA and other proteins associated with spliceosome assembly (Costanzo et al., 2010; Gottschalk et al., 1998; Schwer and Shuman, 2014, 2015; Wilmes et al., 2008) . My aim in this thesis is to study the contributions of Luc7, an essential 261 aa U1 snRNP subunit while also determining the importance of its zinc finger motifs. I further aim to determine the gamut of mutational synergies of benign mutations in Luc7 with other components of the spliceosome. In addition, my thesis also aims to study the contribution of Mud1, an inessential subunit of *S. cerevisiae* U1 snRNP. Mud1 is particularly interesting as it is the homolog of U1A, an essential subunit of human U1

snRNP. Here, I aim to define the functional role of the two RNA recognition motifs (RRM) of Mud1 and the probable role of the linker region connecting the two RRM motifs.

REFERENCES

Dunn E.A., Rader S.D. (2014) Pre-mRNA Splicing and the Spliceosome: Assembly, Catalysis, and Fidelity. In: Sesma A., von der Haar T. (eds) Fungal RNA Biology. Springer, Cham.

Abovich, N., Liao, X.C., Rosbash, M., 1994. The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. Genes Dev 8, 843-854.

Abovich, N., Rosbash, M., 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell 89, 403-412.

Ares, M., Jr., Grate, L., Pauling, M.H., 1999. A handful of intron-containing genes produces the lion's share of yeast mRNA. RNA 5, 1138-1139.

Bai, R., Wan, R., Yan, C., Lei, J., Shi, Y., 2018. Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation. Science 360, 1423-1429.

Bartels, C., Urlaub, H., Luhrmann, R., Fabrizio, P., 2003. Mutagenesis suggests several roles of Snu114p in pre-mRNA splicing. J Biol Chem 278, 28324-28334.

Berget, S.M., Moore, C., Sharp, P.A., 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A 74, 3171-3175.

Bon, E., Casaregola, S., Blandin, G., Llorente, B., Neuveglise, C., Munsterkotter, M., Guldener, U., Mewes, H.W., Van Helden, J., Dujon, B., Gaillardin, C., 2003. Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. Nucleic Acids Res 31, 1121-1135.

Boon, K.L., Auchynnikava, T., Edwalds-Gilbert, G., Barrass, J.D., Droop, A.P., Dez, C., Beggs, J.D., 2006. Yeast ntr1/spp382 mediates prp43 function in postspliceosomes. Mol Cell Biol 26, 6016-6023.

Boone, C., Bussey, H., Andrews, B.J., 2007. Exploring genetic interactions and networks with yeast. Nat Rev Genet 8, 437-449.

Breathnach, R., Chambon, P., 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem 50, 349-383.

Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., Luhrmann, R., 1984. Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. EMBO J 3, 1357-1363.

Brow, D.A., Guthrie, C., 1988. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. Nature 334, 213-218.

Brow, D.A., Guthrie, C., 1989. Splicing a spliceosomal RNA. Nature 337, 14-15.

Brys, A., Schwer, B., 1996. Requirement for SLU7 in yeast pre-mRNA splicing is dictated by the distance between the branchpoint and the 3' splice site. RNA 2, 707-717.

Chan, S.P., Cheng, S.C., 2005. The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. Journal of Biological Chemistry 280, 31190-31199.

Chan, S.P., Kao, D.I., Tsai, W.Y., Cheng, S.C., 2003. The Prp19p-associated complex in spliceosome activation. Science 302, 279-282.

Chapman, K.B., Boeke, J.D., 1991. Isolation and characterization of the gene encoding yeast debranching enzyme. Cell 65, 483-492.

Chen, H.C., Cheng, S.C., 2012. Functional roles of protein splicing factors. Biosci Rep 32, 345-359.

Chen, H.C., Tseng, C.K., Tsai, R.T., Chung, C.S., Cheng, S.C., 2013. Link of NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22. Mol Cell Biol 33, 514-525.

Chen, J.Y., Stands, L., Staley, J.P., Jackups, R.R., Jr., Latus, L.J., Chang, T.H., 2001. Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Mol Cell 7, 227-232.

Cheng, S.C., Abelson, J., 1987. Spliceosome assembly in yeast. Genes Dev 1, 1014-1027.

Chow, L.T., Gelinas, R.E., Broker, T.R., Roberts, R.J., 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell 12, 1-8.

Colot, H.V., Stutz, F., Rosbash, M., 1996. The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes Dev 10, 1699-1708.

Cordin, O., Beggs, J.D., 2013. RNA helicases in splicing. RNA Biol 10, 83-95.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., Prinz, J., St Onge, R.P., VanderSluis, B., Makhnevych, T., Vizeacoumar, F.J., Alizadeh, S., Bahr, S., Brost, R.L., Chen, Y., Cokol, M., Deshpande, R., Li, Z., Lin, Z.Y., Liang, W., Marback, M., Paw, J., San Luis, B.J., Shuteriqi, E., Tong, A.H., van Dyk, N., Wallace, I.M., Whitney, J.A., Weirauch, M.T., Zhong, G., Zhu, H., Houry, W.A., Brudno, M., Ragibizadeh, S., Papp, B., Pal, C., Roth, F.P., Giaever, G., Nislow, C., Troyanskaya, O.G., Bussey, H., Bader, G.D., Gingras, A.C., Morris, Q.D., Kim, P.M., Kaiser, C.A., Myers, C.L., Andrews, B.J., Boone, C., 2010. The genetic landscape of a cell. Science 327, 425-431.

Crawford, D.J., Hoskins, A.A., Friedman, L.J., Gelles, J., Moore, M.J., 2013. Single-molecule colocalization FRET evidence that spliceosome activation precedes stable approach of 5' splice site and branch site. Proc Natl Acad Sci U S A 110, 6783-6788.

Darnell, J.E., Philipson, L., Wall, R., Adesnik, M., 1971. Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA. Science 174, 507-510.

Del Campo, M., Lambowitz, A.M., 2009. Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. Mol Cell 35, 598-609.

Dix, I., Russell, C.S., O'Keefe, R.T., Newman, A.J., Beggs, J.D., 1998. Protein-RNA interactions in the U5 snRNP of Saccharomyces cerevisiae. RNA 4, 1675-1686.

Edmonds, M., Vaughan, M.H., Jr., Nakazato, H., 1971. Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa

cells: possible evidence for a precursor relationship. Proc Natl Acad Sci U S A 68, 1336-1340.

Fabrizio, P., Abelson, J., 1992. Thiophosphates in yeast U6 snRNA specifically affect pre-mRNA splicing in vitro. Nucleic Acids Res 20, 3659-3664.

Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W.S., Luhrmann, R., 1997. An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. EMBO J 16, 4092-4106.

Fortes, P., Kufel, J., Fornerod, M., Polycarpou-Schwarz, M., Lafontaine, D., Tollervey, D., Mattaj, I.W., 1999. Genetic and physical interactions involving the yeast nuclear cap-binding complex. Mol Cell Biol 19, 6543-6553.

Gordon, P.M., Sontheimer, E.J., Piccirilli, J.A., 2000. Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. RNA 6, 199-205.

Gottschalk, A., Tang, J., Puig, O., Salgado, J., Neubauer, G., Colot, H.V., Mann, M., Seraphin, B., Rosbash, M., Luhrmann, R., Fabrizio, P., 1998. A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. RNA 4, 374-393.

Guthrie, C., Patterson, B., 1988. Spliceosomal snRNAs. Annu Rev Genet 22, 387-419.

Hage, R., Tung, L., Du, H., Stands, L., Rosbash, M., Chang, T.H., 2009. A targeted bypass screen identifies Ynl187p, Prp42p, Snu71p, and Cbp80p for stable U1 snRNP/Pre-mRNA interaction. Mol Cell Biol 29, 3941-3952.

Hartwell, L.H., 1967. Macromolecule synthesis in temperature-sensitive mutants of yeast. J Bacteriol 93, 1662-1670.

Hartwell, L.H., McLaughlin, C.S., Warner, J.R., 1970. Identification of ten genes that control ribosome formation in yeast. Mol Gen Genet 109, 42-56.

Hausmann, S., Zheng, S., Costanzo, M., Brost, R.L., Garcin, D., Boone, C., Shuman, S., Schwer, B., 2008. Genetic and biochemical analysis of yeast and human cap trimethylguanosine synthase: functional overlap of 2,2,7-trimethylguanosine caps,

small nuclear ribonucleoprotein components, pre-mRNA splicing factors, and RNA decay pathways. J Biol Chem 283, 31706-31718.

Horowitz, D.S., Abelson, J., 1993. Stages in the second reaction of pre-mRNA splicing: the final step is ATP independent. Genes Dev 7, 320-329.

Hossain, M.A., Johnson, T.L., 2014. Using yeast genetics to study splicing mechanisms. Methods Mol Biol 1126, 285-298.

Izaurralde, E., Lewis, J., Gamberi, C., Jarmolowski, A., McGuigan, C., Mattaj, I.W., 1995. A cap-binding protein complex mediating U snRNA export. Nature 376, 709-712.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., Mattaj, I.W., 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell 78, 657-668.

Jacewicz, A., Schwer, B., Smith, P., Shuman, S., 2014. Crystal structure, mutational analysis and RNA-dependent ATPase activity of the yeast DEAD-box pre-mRNA splicing factor Prp28. Nucleic Acids Res 42, 12885-12898.

Kao, H.Y., Siliciano, P.G., 1996. Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. Mol Cell Biol 16, 960-967.

Kistler, A.L., Guthrie, C., 2001. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. Genes Dev 15, 42-49.

Konarska, M.M., Sharp, P.A., 1987. Interactions between Small Nuclear Ribonucleoprotein-Particles in Formation of Spliceosomes. Cell 49, 763-774.

Kretzner, L., Krol, A., Rosbash, M., 1990. Saccharomyces cerevisiae U1 small nuclear RNA secondary structure contains both universal and yeast-specific domains. Proc Natl Acad Sci U S A 87, 851-855.

Lardelli, R.M., Thompson, J.X., Yates, J.R., 3rd, Stevens, S.W., 2010. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA 16, 516-528.

Lee, C., Jaladat, Y., Mohammadi, A., Sharifi, A., Geisler, S., Valadkhan, S., 2010. Metal binding and substrate positioning by evolutionarily invariant U6 sequences in catalytically active protein-free snRNAs. RNA 16, 2226-2238.

Lewis, J.D., Gorlich, D., Mattaj, I.W., 1996. A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. Nucleic Acids Res 24, 3332-3336.

Li, X., Liu, S., Jiang, J., Zhang, L., Espinosa, S., Hill, R.C., Hansen, K.C., Zhou, Z.H., Zhao, R., 2017. CryoEM structure of Saccharomyces cerevisiae U1 snRNP offers insight into alternative splicing. Nat Commun 8, 1035.

Liao, X.C., Tang, J., Rosbash, M., 1993. An enhancer screen identifies a gene that encodes the yeast U1 snRNP A protein: implications for snRNP protein function in pre-mRNA splicing. Genes Dev 7, 419-428.

Liao, X.L., Kretzner, L., Seraphin, B., Rosbash, M., 1990. Universally conserved and yeast-specific U1 snRNA sequences are important but not essential for U1 snRNP function. Genes Dev 4, 1766-1774.

Libri, D., Graziani, N., Saguez, C., Boulay, J., 2001. Multiple roles for the yeast SUB2/yUAP56 gene in splicing. Genes Dev 15, 36-41.

Linder, P., Fuller-Pace, F.V., 2013. Looking back on the birth of DEAD-box RNA helicases. Biochim Biophys Acta 1829, 750-755.

Liu, H.L., Cheng, S.C., 2012. The interaction of Prp2 with a defined region of the intron is required for the first splicing reaction. Mol Cell Biol 32, 5056-5066.

Lockhart, S.R., Rymond, B.C., 1994. Commitment of yeast pre-mRNA to the splicing pathway requires a novel U1 small nuclear ribonucleoprotein polypeptide, Prp39p. Mol Cell Biol 14, 3623-3633.

Lopez, P.J., Seraphin, B., 2000. YIDB: the Yeast Intron DataBase. Nucleic Acids Res 28, 85-86.

Luhrmann, R., Kastner, B., Bach, M., 1990. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. Biochim Biophys Acta 1087, 265-292.

Lustig, A.J., Lin, R.J., Abelson, J., 1986. The yeast RNA gene products are essential for mRNA splicing in vitro. Cell 47, 953-963.

Madhani, H.D., Guthrie, C., 1994. Dynamic RNA-RNA interactions in the spliceosome. Annu Rev Genet 28, 1-26.

Mazza, C., Segref, A., Mattaj, I.W., Cusack, S., 2002. Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. EMBO J 21, 5548-5557.

McPheeters, D.S., Schwer, B., Muhlenkamp, P., 2000. Interaction of the yeast DExH-box RNA helicase prp22p with the 3' splice site during the second step of nuclear pre-mRNA splicing. Nucleic Acids Res 28, 1313-1321.

Mefford, M.A., Staley, J.P., 2009. Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. RNA 15, 1386-1397.

Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J., Bordonne, R., 2002. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. Mol Cell 9, 891-901.

Neuveglise, C., Marck, C., Gaillardin, C., 2011. The intronome of budding yeasts. Cr Biol 334, 662-670.

Newman, A.J., Norman, C., 1992. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. Cell 68, 743-754.

Nielsen, K.H., Staley, J.P., 2012. Spliceosome activation: U4 is the path, stem I is the goal, and Prp8 is the keeper. Let's cheer for the ATPase Brr2! Genes Dev 26, 2461-2467.

Nilsen, T.W., 2003. The spliceosome: the most complex macromolecular machine in the cell? Bioessays 25, 1147-1149.

Noble, S.M., Guthrie, C., 1996. Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. Genetics 143, 67-80.

O'Day, C.L., Dalbadie-McFarland, G., Abelson, J., 1996. The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. J Biol Chem 271, 33261-33267.

Ohrt, T., Odenwalder, P., Dannenberg, J., Prior, M., Warkocki, Z., Schmitzova, J., Karaduman, R., Gregor, I., Enderlein, J., Fabrizio, P., Luhrmann, R., 2013. Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. RNA 19, 902-915.

Perry, R.P., Kelley, D.E., 1976. Kinetics of formation of 5' terminal caps in mRNA. Cell 8, 433-442.

Plaschka, C., Lin, P.C., Charenton, C., Nagai, K., 2018. Prespliceosome structure provides insights into spliceosome assembly and regulation. Nature.

Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., Nagai, K., 2009. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature 458, 475-480.

Putnam, A.A., Jankowsky, E., 2013. DEAD-box helicases as integrators of RNA, nucleotide and protein binding. Biochim Biophys Acta 1829, 884-893.

Qiu, Z.R., Chico, L., Chang, J., Shuman, S., Schwer, B., 2012. Genetic interactions of hypomorphic mutations in the m7G cap-binding pocket of yeast nuclear cap binding complex: an essential role for Cbc2 in meiosis via splicing of MER3 pre-mRNA. RNA 18, 1996-2011.

Raker, V.A., Plessel, G., Luhrmann, R., 1996. The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. EMBO J 15, 2256-2269.

Rhode, B.M., Hartmuth, K., Westhof, E., Luhrmann, R., 2006. Proximity of conserved U6 and U2 snRNA elements to the 5' splice site region in activated spliceosomes. EMBO J 25, 2475-2486.

Rosbash, M., Harris, P.K., Woolford, J.L., Jr., Teem, J.L., 1981. The effect of temperature-sensitive RNA mutants on the transcription products from cloned ribosomal protein genes of yeast. Cell 24, 679-686.

Rottman, F., Shatkin, A.J., Perry, R.P., 1974. Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing. Cell 3, 197-199.

Schwer, B., 2008. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Mol Cell 30, 743-754.

Schwer, B., Chang, J., Shuman, S., 2013. Structure-function analysis of the 5' end of yeast U1 snRNA highlights genetic interactions with the Msl5*Mud2 branchpoint-binding complex and other spliceosome assembly factors. Nucleic Acids Res 41, 7485-7500.

Schwer, B., Gross, C.H., 1998. Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J 17, 2086-2094.

Schwer, B., Guthrie, C., 1992. A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. EMBO J 11, 5033-5039.

Schwer, B., Shuman, S., 2014. Structure-function analysis of the Yhc1 subunit of yeast U1 snRNP and genetic interactions of Yhc1 with Mud2, Nam8, Mud1, Tgs1, U1 snRNA, SmD3 and Prp28. Nucleic Acids Res 42, 4697-4711.

Schwer, B., Shuman, S., 2015. Structure-function analysis and genetic interactions of the Yhc1, SmD3, SmB, and Snp1 subunits of yeast U1 snRNP and genetic interactions of SmD3 with U2 snRNP subunit Lea1. RNA 21, 1173-1186.

Shi, Y., 2017. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat Rev Mol Cell Biol 18, 655-670.

Shuman, S., 2001. The mRNA capping apparatus as drug target and guide to eukaryotic phylogeny. Cold Spring Harb Sym 66, 301-312.

Shuman, S., 2007. Transcriptional networking cap-tures the 7SK RNA 5'-gamma-methyltransferase. Mol Cell 27, 517-519.

Siliciano, P.G., Kivens, W.J., Guthrie, C., 1991. More than half of yeast U1 snRNA is dispensable for growth. Nucleic Acids Res 19, 6367-6372.

Small, E.C., Leggett, S.R., Winans, A.A., Staley, J.P., 2006. The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. Mol Cell 23, 389-399.

Smith, V., Barrell, B.G., 1991. Cloning of a yeast U1 snRNP 70K protein homologue: functional conservation of an RNA-binding domain between humans and yeast. EMBO J 10, 2627-2634.

Spingola, M., Grate, L., Haussler, D., Ares, M., Jr., 1999. Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. RNA 5, 221-234.

Staley, J.P., Guthrie, C., 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Mol Cell 3, 55-64.

Steitz, T.A., Steitz, J.A., 1993. A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci U S A 90, 6498-6502.

Stevens, S.W., Barta, I., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., Abelson, J., 2001. Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. RNA 7, 1543-1553.

Strauss, E.J., Guthrie, C., 1991. A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. Genes Dev 5, 629-641.

Tang, J., Abovich, N., Fleming, M.L., Seraphin, B., Rosbash, M., 1997. Identification and characterization of a yeast homolog of U1 snRNP-specific protein C. EMBO J 16, 4082-4091.

Teigelkamp, S., Newman, A.J., Beggs, J.D., 1995. Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. EMBO J 14, 2602-2612.

Tsai, R.T., Fu, R.H., Yeh, F.L., Tseng, C.K., Lin, Y.C., Huang, Y.H., Cheng, S.C., 2005. Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev 19, 2991-3003.

Tseng, C.K., Liu, H.L., Cheng, S.C., 2011. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA 17, 145-154.

Venter, J.C., Adams, M.D., Myers, E.W., …Zhu, X., 2001. The sequence of the human genome. Science 291, 1304-1351.

Will, C.L., Luhrmann, R., 2011. Spliceosome structure and function. Cold Spring Harb Perspect Biol 3.

Wilmes, G.M., Bergkessel, M., Bandyopadhyay, S., Shales, M., Braberg, H., Cagney, G., Collins, S.R., Whitworth, G.B., Kress, T.L., Weissman, J.S., Ideker, T., Guthrie, C., Krogan, N.J., 2008. A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. Mol Cell 32, 735-746.

Yean, S.L., Wuenschell, G., Termini, J., Lin, R.J., 2000. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. Nature 408, 881-884.

Zhang, M., Green, M.R., 2001. Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. Genes Dev 15, 30-35.

Chapter 2

Structure–function analysis and genetic interactions of the Luc7 subunit of the *Saccharomyces cerevisiae* U1 snRNP*

INTRODUCTION

pre-mRNA splicing initiates when the U1 snRNP engages the intron 5′ splice site (5′SS). The *Saccharomyces cerevisiae* U1 snRNP consists of a trimethylguanosine (TMG) capped 568-nt U1 snRNA, a seven-subunit Sm protein ring (present also in the U2, U4, and U5 snRNPs), and ten U1-specific protein subunits: Prp39, Prp40, Snu71, Snu56, Snp1, Mud1, Luc7, Prp42, Nam8, and Yhc1(Fortes et al., 1999a; Gottschalk et al., 1998; Schwer et al., 2011). Base-pairing of the U1 snRNA leader motif 5′-ACUUAC sequence with the consensus yeast 5′SS element 5′-GUAUGU nucleates an initial U1•pre-mRNA complex. Cross-intron bridging interactions between the yeast U1 snRNP at the 5′ SS and the Msl5•Mud2 heterodimer at the branchpoint sequence 5′-UACUAAC then stabilize a commitment complex, which provides a scaffold for recruitment of the U2 snRNP to the branchpoint (Abovich and Rosbash, 1997). The U1 snRNP is eventually ejected from the spliceosome at the step when the U5•U4•U6 tri-snRNP complex joins en route to forming a pre-mRNA•U2•U5•U6 spliceosome. Dissociation of U1 snRNP is triggered by the DEAD-box ATPase Prp28 (Chen et al., 2001; Jacewicz et al., 2014; Staley and Guthrie, 1999) acting to disrupt the short U1:5′SS RNA duplex or remodel protein–RNA contacts at the 5′SS (or both).

Traditional and systems genetic approaches, as well as structure-guided mutagenesis, have highlighted a rich network of genetically buffered functions during

* Agarwal, R., Schwer, B., & Shuman, S., 2016. Structure-function analysis and genetic interactions of the Luc7 subunit of the Saccharomyces cerevisiae U1 snRNP. RNA 229, 1302-1310.

early spliceosome assembly in budding yeast, embracing the U1-specific snRNP proteins Mud1, Nam8, Yhc1 and Snp1, the U1 snRNA, the TMG cap, the Cbc2•Sto1 nuclear m7G cap-binding complex (CBC), the DEAD-box ATPase Prp28, and the Msl5•Mud2 branchpoint-binding complex (Abovich et al., 1994; Chang et al., 2012; Colot et al., 1996; Costanzo et al., 2010; Gottschalk et al., 1998; Hausmann et al., 2008; Jacewicz et al., 2015; Liao et al., 1993; Qiu et al., 2012; Schwer et al., 2013; Schwer and Shuman, 2014, 2015; Wilmes et al., 2008). This network is defined by the numerous instances in which null alleles of inessential players (e.g., Mud1, Nam8, Mud2) or benign mutations in essential factors (e.g., Yhc1, Snp1, Msl5) elicit synthetic lethal and sick phenotypes when combined with other benign mutations in the splicing machinery.

Several of the essential subunits of the *S. cerevisiae* U1 snRNP are still comparatively uncharted with respect to their structure–activity relations and genetic interactions. In the present study, I focus on Luc7, a 261-aa polypeptide encoded by the YDL087C ORF on chromosome 4. The *LUC7* gene was identified initially in a genetic screen for synthetic lethality with *cbc2Δ sto1Δ* double-deletion of the nuclear CBC subunits (Fortes et al., 1999b).The "LUC" screen (**L**ethal **U**nless **C**BC is produced) yielded 14 complementation groups of *LUC* mutants, many of which could be assigned to known spliceosome components, including (i) U1 snRNP subunits Mud1, Nam8, Snu56, and Snu71; (ii) the SmD3 subunit of the Sm ring; and (iii) the Mud2 subunit of the branchpoint binding protein (Fortes et al., 1999b). Subsequent studies showed that Luc7 is a stoichiometric subunit of the yeast U1 snRNP and that a *luc7-ts* allele was associated with reduced efficiency of splicing of a reporter pre-mRNA with a nonconsensus 5′SS sequence GUAUAU (Fortes et al., 1999a). U1 snRNP isolated from a *luc7-ts* strain grown at restrictive temperature was deficient in

both Luc7 and Snu71 subunits (Fortes et al., 1999a). Two-hybrid experiments identified a physical interaction between Luc7 and the essential FF1 domain of Prp40 (Ester and Uetz, 2008). It was suggested that the tandem FF domains of Prp40 nucleate a Prp40•Luc7•Snu71 subcomplex with the yeast U1 snRNP (Ester and Uetz, 2008). In vitro crosslinking of Luc7 to the 5′ exon of a ^{32}P-labeled 4-thio-U substituted pre-mRNA substrate that was incubated in a yeast extract, in a manner dependent on the 5′SS consensus sequence in the pre-mRNA, led to the suggestion that Luc7 stabilizes the U1 snRNP•pre-mRNA complex (Puig et al., 2007).

The primary structure of *S. cerevisiae* Luc7 is shown in Figure 2.1, aligned to the Luc7 homologs from *Aspergillus fumigatus* (Genbank accession XP_750789), *Trichoderma atroviride* (Genbank accession EHK44709), and *Homo sapiens* (Luc7-like 1 isoform a; Genbank accession NP_060502). As noted previously (Puig et al., 2007), Luc7 proteins are distinguished by two putative zinc finger (ZnF) motifs, which I refer to henceforth as ZnF1 (comprising Cys45, Cys53, Cys68, and His72 in *S. cerevisiae* Luc7) and ZnF2 (Cys201, Cys204, His220, and His226 in *S. cerevisiae* Luc7) (Fig. 2.1). To our inspection, there are no other instructive amino acid sequence motifs in *S. cerevisiae* Luc7. The longer 325-amino acid human Luc7-like polypeptide has a unique C-terminal segment that is rich in runs of arginines, with interspersed acidic residues, and Ser-Arg dipeptide repeats (Fig. 2.1). The *A. fumigatus* and *T. atroviride* Luc7 proteins have C-terminal extensions peppered with runs of glycines and Arg-Gly repeats, respectively. *S. cerevisiae* Luc7 has no such low-complexity C-terminal segment. A key distinction between the *S. cerevisiae* and human Luc7 proteins is that the former is an integral subunit of the yeast U1 snRNP, whereas the latter is not a subunit of the human U1 snRNP, which contains only three U1-specific

subunits: U1-70K, U1-A, and U1-C, which are the human homologs of yeast Snp1, Mud1, and Yhc1, respectively (Kondo et al., 2015).

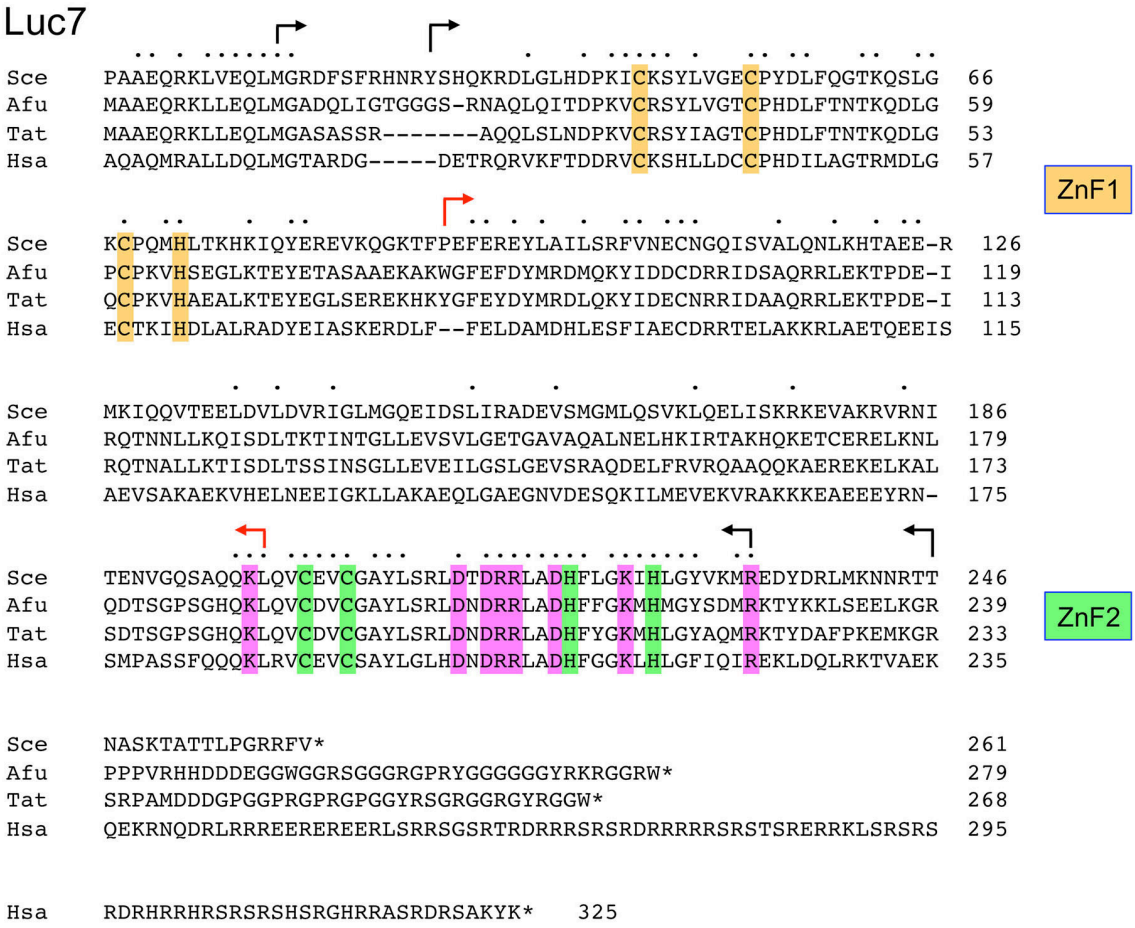


Figure 2.1 Phylogeny-guided mutational analysis of Luc7. The primary structures of the Luc7 polypeptides of *Saccharomyces cerevisiae* (Sce), *Aspergillus fumigatus* (Afu), *Trichoderma atroviride* (Tat), and *Homo sapiens* (Has) are aligned. Positions of side chain identity/similarity are indicated by •. Gaps in the alignment are denoted by dashes. Forward and reverse arrowheads indicate the boundaries of *S. cerevisiae* Luc7 N-terminal and C-terminal truncations, respectively. Black arrowheads denote biologically active truncations; red arrows indicate lethal truncations. The conserved amino acids of the N-terminal CCCH ZnF1 module are shaded gold; the amino acids of the conserved C-terminal CCHH ZnF2 module are shaded green. Eight other conserved positions subjected to mutagenesis in the present study are shaded magenta.

Here I present a phylogenetically guided *in vivo* mutational analysis of S. *cerevisiae* Luc7, entailing N- and C-terminal truncations and alanine scanning of conserved amino acids. I test a series of 23 alleles for phenotypes and for synthetic genetic interactions with a broad spectrum of other early-stage spliceosome components or splicing factors. I define Luc7-(31–246) as a minimal functional unit and demonstrate that whereas mutations of the CCHH ZnF2 motif are lethal, mutations of the ZnF1 CCCH motif are not. A hierarchy of mutational synergies of otherwise benign Luc7 alleles highlights the importance of the N-terminal 18-aa segment and amino acids surrounding ZnF2. By identifying *LUC7* alleles that bypass the essentiality of Prp28, I implicate specific constituents of Luc7 in stabilizing the U1•5′SS interaction.

EXPERIMENTAL PROCEDURES

Luc7 expression plasmids and mutants

A 1.8-kb DNA segment bearing the *LUC7* gene was amplified from *S. cerevisiae* genomic DNA by PCR using a forward primer that introduced a SacI site 500-bp upstream of the Luc7 translation initiation site and a reverse primer to introduce a XhoI site 500-bp downstream from the stop codon. The PCR product was digested with SacI and XhoI and inserted between the SacI and XhoI sites of pRS316 *(CEN URA3)*. A second yeast expression plasmid, p415-LUC7 *(CEN LEU2)*, containing the *LUC7* ORF plus 500 bp of upstream and 500 bp of downstream genomic DNA, was constructed so as to introduce a 5′ BamHI site and a 3′ HindIII site immediately flanking the ORF. N-terminal truncation Luc7-(19–261) was generated by PCR using a forward primer that introduced a BamHI site preceding the Met19 codon. Luc7-(31–261) was generated with a forward primer that introduced a 5′ BamHI site and a Met codon in lieu of the Tyr30 codon. C-terminal truncations were

generated by PCR using reverse primers that introduced stop codons and flanking HindIII sites in lieu of Asn106, Gln199, Glu234 or Asn247. The truncated PCR products were inserted into pRS415-LUC7 in lieu of the wild-type ORF to generate plasmids p415-LUC7-(19–261), p415-LUC7-(31–261), p415-LUC7-(1–198), p415-LUC7-(1–233), p415-LUC7-(1–246), and p415-LUC7-(31–246). Single alanine mutations were introduced into the *LUC7* gene by two-stage PCR overlap extension with mutagenic primers. The PCR products were digested with BamHI and HindIII and inserted into the pRS415-based expression vector to generate a series of p415-LUC7-Ala plasmids. The *LUC7* genes were sequenced completely to confirm that no unwanted changes were acquired during amplification and cloning.

Yeast strains and tests of Luc7 function in vivo

To develop a plasmid shuffle assay for gauging mutational effects on Luc7 function, I generated a *luc7Δ* strain that relies for viability on maintenance of a *CEN URA3 LUC7* plasmid (p316-LUC7). A BY4743 diploid strain in which one chromosomal *LUC7* locus from ORF positions +1 to +787 was replaced with a *kanMX* cassette was transformed with p316-LUC7. The diploid was sporulated, asci were dissected, and haploid *luc7Δ* Ura+ progeny were recovered. The *luc7Δ* [p316-LUC7] cells were resistant to kanamycin and unable to grow on medium containing 0.75 mg/mL FOA (5-fluoroorotic acid). The *luc7Δ* [p316-LUC7] cells were transfected with *CEN LEU2 LUC7* plasmids expressing wild-type or mutated Luc7 proteins. Individual $Leu2^+$ transformants were selected and streaked on agar medium containing FOA. *LUC7* mutant alleles that failed to give rise to FOA-resistant colonies at any temperature after incubation for 8 d at 20°C, 30°C, or 37°C were deemed lethal. Individual FOA resistant colonies with viable *LUC7* alleles were grown to midlog phase in YPD broth and adjusted to A_{600} of 0.1. Aliquots (3 μL) of serial 10-fold

dilutions were spotted to YPD agar plates, which were incubated at 20°C, 25°C, 30°C, 34°C, and 37°C.

Tests of mutational synergy

I developed plasmid shuffle assays to test the effects of Luc7 mutations in *mud2Δ, nam8Δ, mud1Δ, tgs1Δ, cbc2-Y24A, prp40-ΔWW,* and *snp1-(1–223)* genetic backgrounds. Via standard genetic manipulations of mating, sporulation, and dissection, I generated *mud2Δ luc7Δ* [p316-LUC7], *nam8Δ luc7Δ* [p316-LUC7], *mud1Δ luc7Δ* [p316-LUC7], *tgs1Δ luc7Δ* [p316-LUC7], *cbc2-Y24A luc7Δ* [p316-LUC7], *prp40-ΔWW luc7Δ* [p316-LUC7], and *snp1Δ luc7Δ* [p316-SNP1-LUC7] haploid strains, which were unable to grow on FOA-containing medium unless they had been previously transformed with wild-type *LUC7* or a functional *LUC7* mutant allele on a *CEN LEU2* plasmid or, in the case of *snp1Δ luc7Δ* [p316-SNP1- LUC7], a *CEN LEU2 LUC7* plasmid plus a *CEN HIS3* plasmid bearing the *snp1-(1–223)* allele.

Test of Prp28 bypass

Heterozygous *prp28Δ/PRP28 luc7Δ/LUC7* diploids were obtained by crossing *prp28Δ::natMX* p316-PRP28 (CEN URA3 PRP28) with *luc7Δ::kanMX* p316-LUC7 cells of the opposite mating type, selecting diploids on YPD medium containing clonNat plus G418, and plating them to FOA-containing medium to select against the two *URA3* plasmids. The heterozygous diploid was then transfected with a *CEN URA3 LUC7 PRP28* plasmid (p316-LUC7-PRP28) in which the *LUC7* gene (nucleotides −500 to +500) is arranged in a head-to-tail configuration with the *PRP28* gene (nucleotides −520 to +560). Ura+ heterozygous diploids were subjected to sporulation and tetrad dissection, after which haploid *luc7Δ prp28Δ* [p316- LUC7-PRP28] progeny were recovered. *The luc7Δ prp28Δ* [p316- LUC7-PRP28] strain was unable to

grow on FOA medium unless it had been subjected to prior cotransformation with *CEN LEU2 LUC7* and *CEN HIS3 PRP28* plasmids bearing the wild-type *LUC7* and *PRP28* genes. Transformation with the *CEN LEU2 LUC7* plasmid alone yielded no FOA-resistant survivors. To test for bypass suppression of *prp28Δ* by *LUC7* mutants, *luc7Δ prp28Δ* [p316-LUC7-PRP28] cells were transfected with *CEN LEU2* plasmids expressing truncated or alanine-substituted Luc7 proteins. Individual Leu2+ transformants were selected and streaked on agar medium containing FOA. *LUC7* mutant alleles that gave rise to FOA-resistant colonies after incubation for up to 8 d at 20°C, 30°C, or 37°C were deemed Prp28 bypass suppressors. Viable *prp28Δ* colonies with bypass *LUC7* alleles were grown to mid-log phase in YPD broth and adjusted to A_{600} of 0.1. Aliquots (3 μL) of serial 10-fold dilutions were spotted to YPD agar plates, which were incubated at 20°C, 25°C, 30°C, 34°C, and 37°C.

RT-PCR assays of pre-mRNA splicing in vivo

Yeast *luc7Δ* p316-LUC7 haploid strains with wild-type *LUC7* or a "benign" *LUC7* mutant allele (i.e., no growth defect on YPD agar) on a *CEN LEU2* plasmid were grown in liquid YPD medium at 30°C until A_{600} reached 0.6–0.8. Cells (20 A_{600} units) were harvested by centrifugation and total cellular RNA was isolated by the hot phenol method. The RNA preparations were treated with DNase I. First-strand cDNA synthesis was carried out by using the ProtoScript First Strand cDNA synthesis kit (NEB) with 1 μg of total RNA template and an oligo$(dT)_{23}$ primer. Aliquots of the mixtures were then used for 35 cycles of PCR amplification by Taq polymerase (NEB) of the *SUS1* cDNAs with 0.2 μM oligonucleotide primers 5′-TGGATACTGCGCAATTAAAGAGTC and 5′-TCATTGTGTAT CTACAATCTCTTCAAG, complementary to the first and third exons of the *SUS1* transcript, and of the *GLC7* cDNAs with 0.2 μM primers

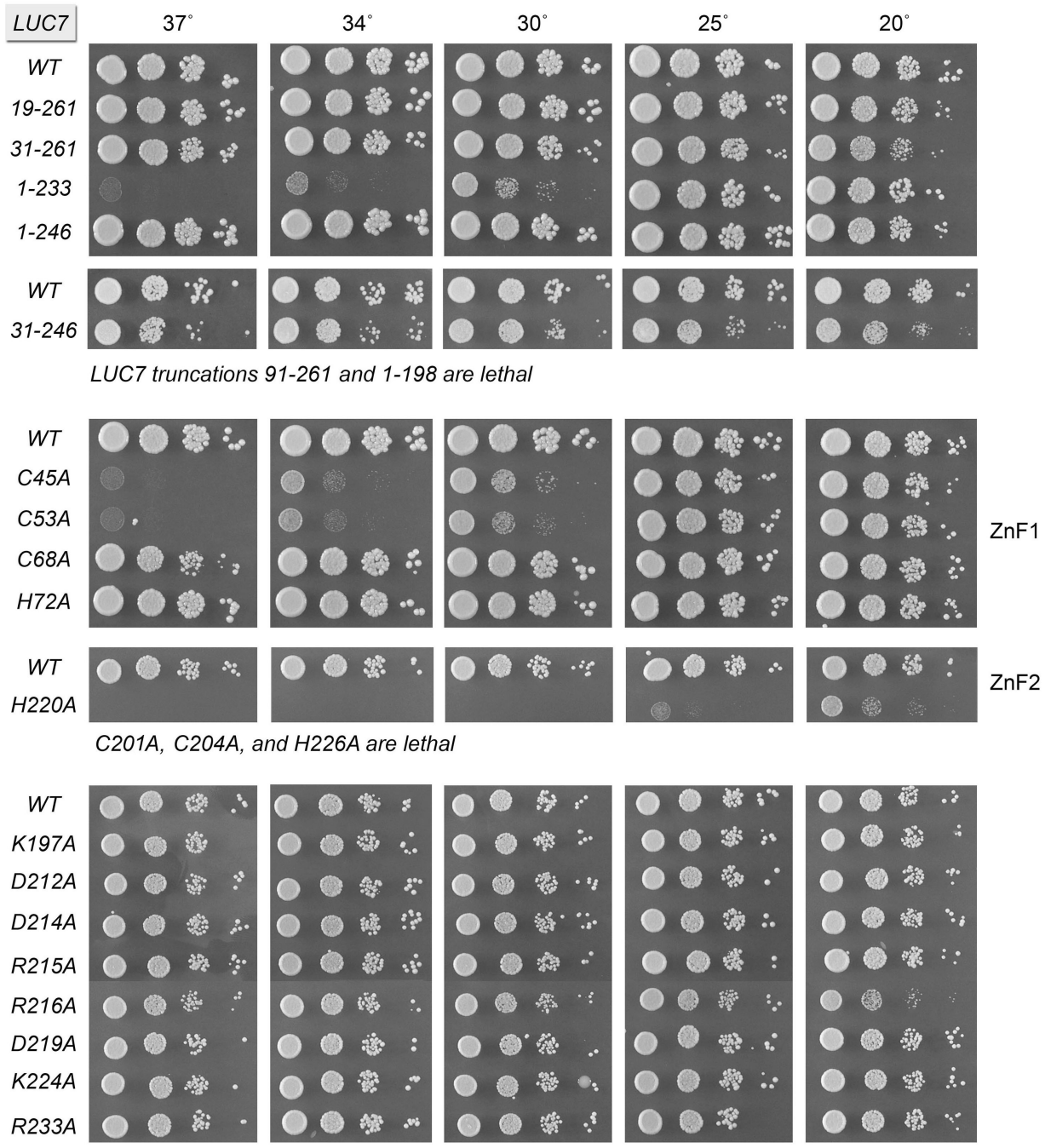
5′-CCTGGTCAACAAGTTGATCTAGAAG and 5′-CTAA GGATTGTTTACCACGGTCGAC, complementary to the two *GLC7* exons. The PCR products were resolved by electrophoresis through a native 2% agarose gel and visualized by staining with ethidium bromide.

RESULTS

Mapping the proximal and distal margins of the functional Luc7 protein

A series of N-terminal and C-terminal truncation alleles were placed on *CEN LEU2* plasmids under the control of the native *LUC7* promoter and tested by plasmid shuffle for complementation of a *luc7Δ* p[*CEN URA3 LUC7*] strain. I found that deleting 18 or 30 amino acids from the N terminus (black arrows in Fig. 2.1) had no apparent impact on yeast growth at 25°C–37°C, as gauged by colony size (Fig. 2.2). (The *LUC7-[31–261]* strain formed smaller colonies at 20° C.) In contrast, an N-terminal truncation of 90 amino acids (red arrow in Fig. 2.1), which eliminated the ZnF1 module, was lethal (i.e., no FOA-resistant colonies were recovered after plasmid shuffle performed at 20°C, 30°C, and 37°C). Whereas deleting the C-terminal segment from amino acids 247–261 had no effect on yeast growth, further deletion of amino acids 234–246 elicited a temperature-sensitive (Gottschalk et al.) growth defect, i.e., the *LUC7-(1–233)* strain grew as well as wild-type *LUC7* at 20°C–25°C but grew slowly at 30°C and failed to grow at 34°C–37°C (Fig. 2.2). The more extensive C-terminal deletion *LUC7-(1–198)*, which eliminated the ZnF2 module, was lethal. To query whether the dispensable N-terminal and C-terminal segments make genetically

Figure 2.2. Deletion and alanine-scanning mutagenesis of yeast Luc7. Truncated *LUC7* alleles (*top panel*), alanine-mutants of the cysteines and histidines of the ZnF1 and ZnF2 motifs (*middle panels*), and additional alanine mutants in the ZnF2 domain (*bottom panel*) were tested for activity by plasmid shuffle. The viable FOA-resistant *luc7*Δ strains bearing the indicated *LUC7* alleles were spot-tested for growth on YPD agar at the temperatures specified. *LUC7* alleles listed at the *bottom* of each panel failed to complement *luc7*Δ in the plasmid shuffle assay and were deemed lethal.


LUC7
37˚
34˚
30˚
25˚
20˚
WT
19-261
31-261
1-233
1-246
WT
31-246
LUC7 truncations 91-261 and 1-198 are lethal
WT
C45A
C53A
C68A
H72A
ZnF1
WT
H220A
ZnF2
C201A, C204A, and H226A are lethal
WT
K197A
D212A
D214A
R215A
R216A
D219A
K224A
R233A

redundant contributions to Luc7 activity in vivo, I combined the deletions and found that the *LUC7-(31–246)* strain was viable, albeit slightly cold-sensitive (*cs*) (Fig. 2.2).

Alanine scanning mutagenesis of the ZnF1 and ZnF2 cysteines and histidines

Cys45, Cys53, Cys68, and His72 comprising the ZnF1 motif were mutated individually to alanine. The *LUC7 C68A* and *H72A* strains thrived at all temperatures from 20°C–37°C. The *C45A* and *C53A* strains grew well at 20°C–25°C but grew very slowly at 30°C–34°C (as gauged by colony size) and failed to grow at 37°C (Fig. 2.2). Thus, whereas a deletion of the segment from amino acids 36–90 spanning the ZnF1 motif was lethal, elimination of any one of the four putative metal-binding side chains of ZnF1 was not. These results are consistent with scenarios in which (i) the imputed metal-binding function of ZnF1 is not essential for Luc7 bioactivity; (ii) any three of the four ZnF1 amino acids suffices for metal-binding and bioactivity; or (iii) ZnF1 function is genetically redundant with that of ZnF2.

Entirely different results were obtained when I introduced alanines in lieu of the Cys201, Cys204, His220, and His226 residues of the ZnF2 motif. The *C201A*, *C204A*, and *H226A* mutations were lethal. The *H220A* allele gave rise to FOA-resistant colonies at 20°C during the plasmid shuffle procedure; however, the *H220A* strain was barely viable on YPD agar at 20°C and failed to grow at higher temperatures (Fig. 2.2). The fact that each of the imputed metal ligands of ZnF2 is critical for Luc7 bioactivity, whereas those of ZnF1 are not, vitiates the idea that ZnF1 and ZnF2 are functionally redundant.

Alanine scanning of conserved amino acids in the ZnF2 module

Primary structure conservation among the Luc7 proteins aligned in Figure 2.1 is greatest within the segment of *S. cerevisiae* Luc7 from amino acids 196–233 that

embraces ZnF2, i.e., 25/38 positions of side chain identity in all four Luc7 homologs. I extended the alanine scan to eight invariant basic or acidic residues in this ZnF2 module: Lys197, Asp212, Asp214, Arg215, Arg216, Asp219, Lys224, and Arg233 (shaded magenta in Fig. 2.1). All of the eight *LUC7-Ala* strains were viable and grew as well as wild-type *LUC7* on YPD agar at 25°C– 37°C (Fig. 2.2). Only *R216A* cells were slightly slow-growing at 20°C, as gauged by colony size (Fig. 2.2).

Synthetic genetic interactions of Luc7 truncation mutants

To survey genetic interactions of the Luc7 truncation mutants, I tested by plasmid shuffle for complementation of strains in which the genes encoding U1 snRNP subunits Nam8 or Mud1 or the branchpoint binding protein subunit Mud2 were deleted in the *luc7Δ* p[*CEN URA3 LUC7*] background. Yeast *nam8Δ*, *mud1Δ*, and *mud2Δ* single mutants grow as well as wild-type yeast at all temperatures. I also tested for complementation of a *luc7Δ* shuffle strain lacking the gene encoding the TMG capping enzyme Tgs1. Yeast *tgs1Δ* cells grow well at 30°C–37°C, but they fail to grow at 20°C. In addition, I tested complementation of a *luc7Δ* shuffle strain in which the Cbc2 subunit of the nuclear m7G cap-binding complex has a Y24A mutation in the cap binding pocket. Although yeast *cbc2- Y24A* cells grow as well as wild-type yeast at all temperatures, the *Y24A* mutation elicits synthetic lethality or sickness with *nam8Δ, mud1Δ, and mud2Δ* (Qiu et al., 2012). The results are summarized in Figure 2.3.

	mud1Δ	*nam8Δ*	*mud2Δ*	*tgs1Δ*	*cbc2-Y24A*	*prp40-ΔWW*	*snp1(1-223)*
19-261	lethal	lethal	lethal	lethal	lethal	cs	lethal
31-261	lethal	lethal	lethal	lethal	lethal	very sick	lethal
1-246							
1-233	very sick	lethal	lethal	lethal	lethal	very sick	lethal
C45A	lethal	lethal	lethal	lethal	lethal	very sick	lethal
C53A	lethal	lethal	lethal	lethal	lethal	very sick	lethal
C68A	ts	ts	lethal	lethal	lethal		ts
H72A							
K197A			very sick		cs		
D212A	ts, cs	cs	lethal	cs	cs		cs
D214A	very sick	very sick	lethal	very sick	lethal		very sick
R215A	very sick	lethal	lethal	very sick	lethal		lethal
R216A	lethal	lethal	lethal	lethal	lethal	very sick	lethal
D219A	ts, cs	cs	lethal	sick	lethal		cs
K224A			cs		cs		
R233A							

Figure 2.3. Synthetic interactions of Luc7 mutants. Synthetically lethal pairs of alleles are highlighted in red boxes. Other negative pairwise interactions are classified as sick or very sick (yellow boxes) or temperature-sensitive or cold sensitive (*cs*) (light green boxes). Grey boxes denote lack of mutational synergy.

The genetics highlight the importance of the N-terminal segment for Luc7 activity in vivo when other components of the U1 snRNP and the early spliceosome are perturbed. To wit, the *LUC7-(19–261)* allele was lethal in combination with *mud2Δ, nam8Δ, mud1Δ, tgs1Δ*, and *cbc2-Y24A*. The same synthetic lethalities were

observed for the *LUC7-(31–261)* allele (Fig. 2.3). In contrast, deletion of the C-terminal segment in *LUC7-(1–246)* elicited no growth defects in the *mud2Δ, nam8Δ, mud1Δ, tgs1Δ,* or *cbc2-Y24A* backgrounds (Fig. 2.3). However, incremental deletion of the segment from amino acids 234–246 in *LUC7-(1–233),* which resulted in a *ts* growth defect by itself, was synthetically lethal with *mud2Δ, nam8Δ, tgs1Δ*, and *cbc2-Y24A* and synthetically sick with *mud1Δ* (Fig. 2.3).

Unlike the "optional" splicing factors surveyed above, the U1 snRNP subunits Prp40 and Snp1 are essential for yeast vegetative growth. Prp40 is a 583-aa protein containing two tandem WW modules at the N terminus (amino acids 1–70) and four FF domains dispersed downstream (Ester and Uetz, 2008). Previously, a truncated chromosomal allele, *prp40-(77–583)* (henceforth named *prp40-ΔWW*) was constructed, that cleanly subtracts the tandem WW modules and found that the *prp40-ΔWW* strain grew as well as the *PRP40* control strain on YPD agar at all temperatures (Schwer et al., 2013). Here I tested the Luc7 truncation mutants for complementation of a *prp40-ΔWW luc7Δ* shuffle strain. *LUC7-(19–261)* displayed a *cs* growth defect in the *prp40-ΔWW* background. *LUC7-(31–261)* was synthetically sick at 34°C–37°C in combination with *prp40-ΔWW*; the *LUC7-(31–261) prp40-ΔWW* strain failed to grow at 20°C–30°C (classified as "very sick" in Fig. 2.3). *LUC7-(1–246)* elicited no growth phenotype in the *prp40-ΔWW* background. The *LUC7-(1–233) prp40-ΔWW* strain was sick at 20°C–25°C and failed to thrive at 30°C–37°C (Fig. 2.3, very sick).

Snp1, a 300-aa polypeptide, is the yeast homolog of human U1-70K (Kao and Siliciano, 1992). As reported previously, yeast snp1Δ cells bearing a C-terminal truncation allele, *snp1(1–223)*, grow as well as wild-type SNP1 cells on YPD agar at all temperatures (Qiu et al., 2015). *snp1(1–223)* displays a distinctive spectrum of mutational synergies whereby it is (i) synthetically lethal with *mud2Δ*; (ii) severely

synthetically sick and *cs* with *nam8Δ*; (iii) benign in combination with *mud1Δ*; and (iv) suppresses the *cs* defect of *tgs1Δ* (Qiu et al., 2015). Testing the Luc7 truncation mutants for complementation of a *snp1(1-223) luc7Δ* shuffle strain revealed that the N-terminal Luc7 truncations *LUC7-(19–261)* and *LUC7- (31–261)* and the C-terminal truncation *LUC7-(1–233)* were lethal in combination with *snp1(1–223)*. In contrast, the C-terminal truncation *LUC7-(1–246)* was benign in the *snp1 (1-223)* background (Fig. 2.3).

Synthetic genetic interactions of Luc7-Ala mutants

All of the viable Luc7-Ala mutants were evaluated for mutational synergies in the seven genetic backgrounds listed in Figure 2.3. At the severest end of the synergy spectrum was ZnF2 mutant *R216A*, which was benign on its own but synthetically lethal with *mud1Δ, nam8Δ, mud2Δ, tgs1Δ, cbc2-Y24A*, and *snp1(1–223)* and synthetically very sick with *prp40-ΔWW* (Fig. 2.3). Mutations of ZnF2 residues flanking Arg216 also synergized strongly. The neighboring upstream ZnF2 mutation *R215A* was lethal with *nam8Δ, mud2Δ, cbc2-Y24A,* and *snp1(1–223)* and very sick with *mud1Δ* and *tgs1Δ*. At the next amino acid upstream, D214A was lethal with *mud2Δ* and *cbc2-Y24A* and very sick *with mud1Δ, nam8Δ, tgs1Δ,* and *snp1(1–223)*.Moving two residues further upstream, *D212A* was lethal with *mud2Δ*, elicited *cs* defects in the *nam8Δ*, *cbc2-Y24A*, and *snp1(1–223)* backgrounds, caused *ts* and *cs* defects in *mud1Δ* cells, and exacerbated the cold-sensitive growth defect of *tgs1Δ* cells (i.e., raising the restrictive temperature) (Fig. 2.3). The neighboring downstream mutation *D219A* was lethal with *mud2Δ* and *cbc2-Y24A*, sick with *tgs1Δ*, *cs* with *nam8Δ* and *snp1(1-223)*, and both *cs* and *ts* with *mud1Δ* (Fig. 2.3). The mutations at the outer margins of the conserved ZnF2 module displayed milder synergies that were

confined to *mud2Δ* and *cbc2-Y24A* (e.g., *K197A* and *K224A*) or elicited no synthetic phenotype in any of the backgrounds surveyed (e.g., *R233A*).

Synergies of ZnF1 alanine mutations ranged from the entirely benign in the case of *H72A* to the pan-catastrophic for *C45A* and *C53A*. Whereas *C45A* and *C53A* caused *ts* growth defects by themselves, these alleles were lethal with *mud1Δ*, *nam8Δ*, *mud2Δ*, *tgs1Δ*, *cbc2-Y24A*, and *snp1(1– 223)* and very sick with *prp40-ΔWW*. The *C68A* mutation, which caused no growth phenotype, was lethal with *mud2Δ*, *tgs1Δ*, and *cbc2-Y24A* and *ts* with *mud1Δ*, *nam8Δ*, and *snp1(1–223)* (Fig. 2.3).

Luc7 mutations bypass the essentiality of Prp28

Yeast Prp28 is an essential DEAD-box ATPase implicated in displacing the U1 snRNP from the 5′SS during the transition from a pre-mRNA•U1•U2 spliceosome to a pre-mRNA• U2•U5•U6 spliceosome. The requirement for Prp28 for vegetative growth can be bypassed by mutations in the essential U1 snRNP subunits Yhc1, Prp42, and Snu71, and by specific U1 snRNA mutations located within and flanking the segment that base-pairs with the intron 5′SS (Chen et al., 2001; Hage et al., 2009; Schwer et al., 2013; Schwer and Shuman, 2014, 2015) the common theme being that such mutations are thought to weaken the U1•5′SS contacts and thereby allay the need for Prp28 during U1 snRNP ejection from the early spliceosome.

Here I queried whether any of the Luc7 mutations might bypass Prp28, by transforming a *prp28Δ luc7Δ* p[*CEN URA3 PRP28 LUC7*] strain with *CEN LEU2 LUC7* plasmids and selecting for growth on medium containing FOA. Control experiments affirmed that plasmid shuffle with a wild-type *LUC7* plasmid yielded no FOA-resistant survivors, whereas cotransformation with a wild-type *LUC7* plasmid and a *CEN HIS3 PRP28* plasmid did. I found that the N-terminal truncations *LUC7-*

(19–261) and *LUC7-(31–261)* were able to sustain growth of *prp28Δ luc7Δ* cells on FOA, as were several of the *LUC7-Ala* mutants. The viable *luc7Δ prp28Δ* cells bearing these bypass alleles were tested for growth on YPD agar (Fig. 2.4). *LUC7-(19–261) prp28Δ* cells grew as well as wildtype *LUC7 PRP28* cells at 37°C, 34°C, and 30°C, as gauged by colony size, but they grew slowly at 25°C and did not thrive at 20°C. *LUC7-(31–261)* prp28Δ cells grew as well as wildtype at 37°C and 34°C, were slower growing at 30°C and 25°C, and did not thrive at 20°C.

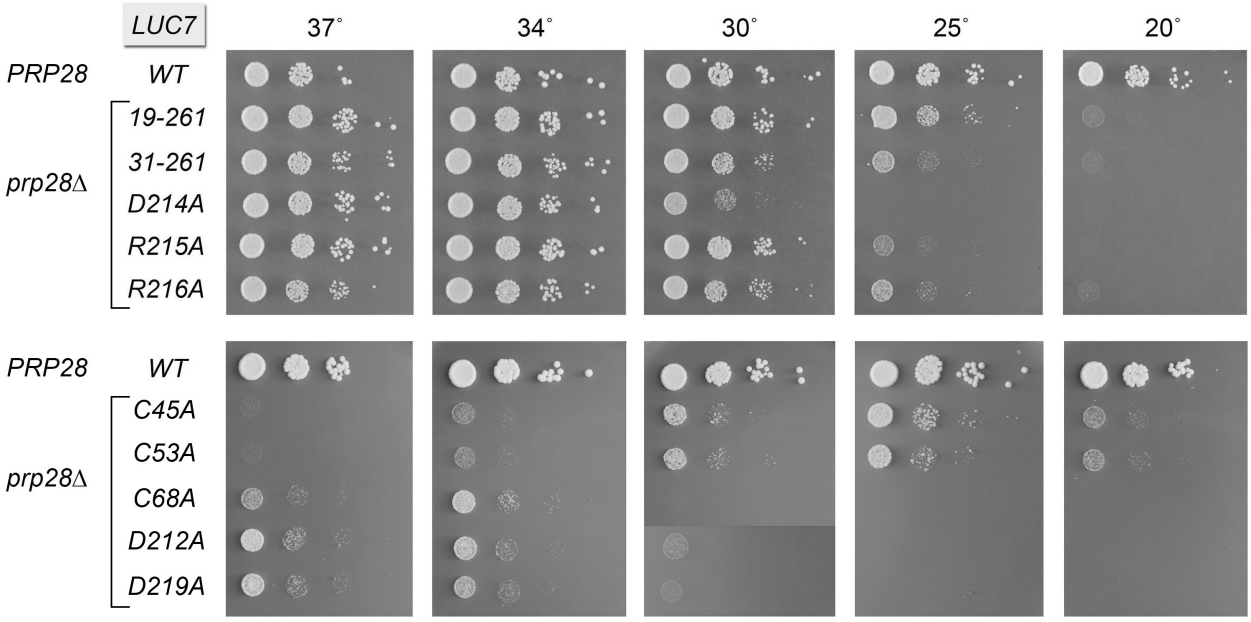


Figure 2.4. Luc7 mutations that bypass Prp28. Yeast *prp28Δ luc7Δ* cells bearing the indicated *LUC7* mutant allele on a *CEN LEU2* plasmid were spot-tested for growth on YPD agar at the temperatures specified. Control strains were those that had been co-transformed with plasmids bearing the wild-type *PRP28* and *LUC7* genes.

Among the Ala mutants, the ZnF2 module allele *R215A* evinced the strongest Prp28 bypass phenotype, i.e., *LUC7-R215A prp28Δ* cells grew well at 37°C, 34°C, and 30°C, but did not thrive at 25°C and 20°C. Flanking mutations *D214A* and *R216A* also bypassed Prp28 to permit wild-type growth at 37°C and 34°C, while sustaining

slow or no growth at lower temperatures. ZnF2 alleles *D212A* and *D219A* were comparatively weak Prp28 bypass suppressors, in that they sustained slow growth of *prp28Δ* cells and only at 34°C–37°C. I surmise that higher temperatures destabilize the weakened U1•5′SS interface in cells bearing the aforementioned Luc7 mutants and thereby permit U1 snRNP ejection without the assistance of Prp28. However, when the U1:5′SS duplex pairing is more stable at lower temperatures, Prp28 is still required. Cold-sensitivity is a characteristic shared with most of the other prp28 bypass mutations described to date (Chen et al., 2001; Hage et al., 2009; Schwer et al., 2013; Schwer and Shuman, 2014, 2015)

Exceptions to the cold-sensitivity of Prp28 bypass were evident for ZnF1 alleles *C45A* and *C53A*, which supported slow growth of *prp28Δ* cells at 25°C but did not allow growth at 37°C (Fig. 2.4). The lack of bypass at 37°C reflects the fact that the *C45A* and *C53A* mutations by themselves elicit a ts growth defect (Fig. 2.2). In contrast, ZnF1 mutant *C68A*, which has no growth phenotype on its own, affords a relatively weak Prp28 bypass, but it does so in the typical *cs* fashion (Fig. 2.4).

Taken together, the mutational synergies and Prp28 bypass experiments underscore the significant contributions of the Luc7 N terminus and the ZnF1 and ZnF2 modules to early spliceosome assembly and stability.

Luc7 mutations affect splicing of SUS1 pre-mRNA

I surveyed our collection of Luc7 mutations that had no effect per se on yeast growth for their effects on the efficacy of in vivo splicing of two specific yeast pre-mRNAs: *GLC7,* which has a single intron and is efficiently spliced in wildtype cells (Fig. 2.5A); and *SUS1*, a two-intron-containing transcript (Fig. 2.5B). I chose

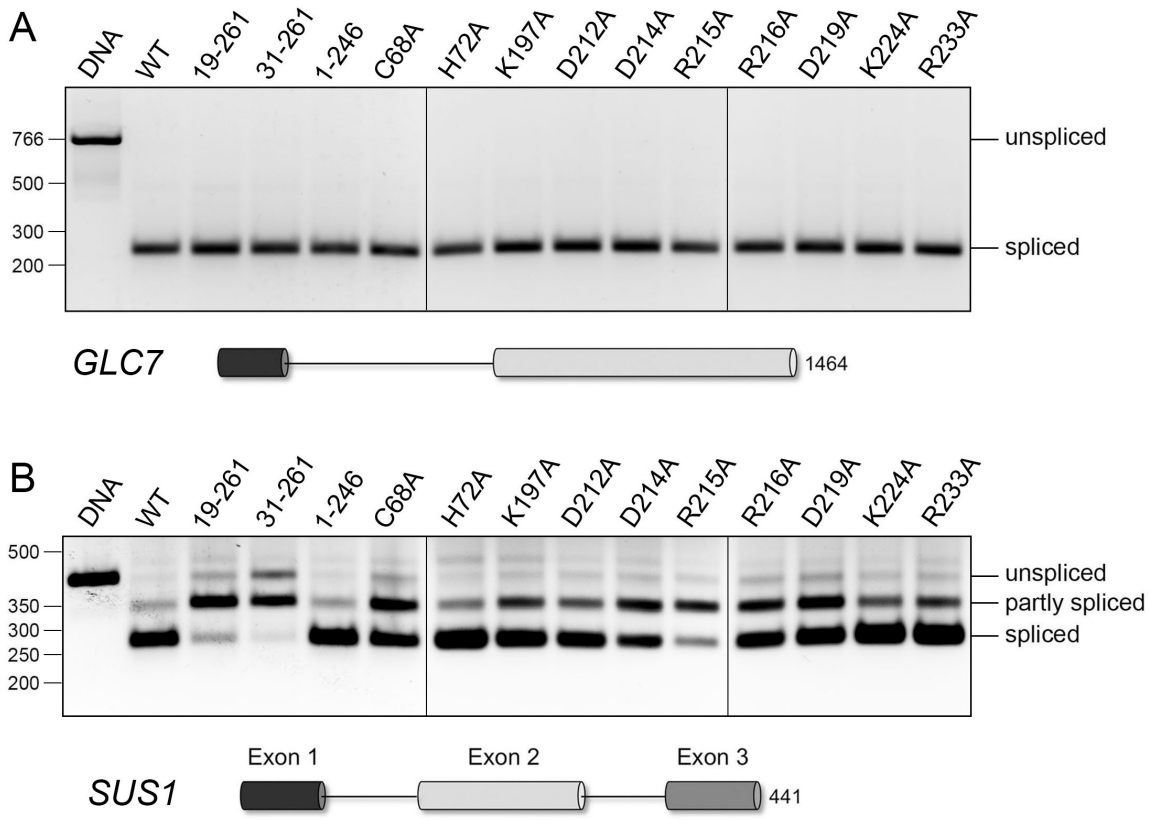


Figure 2.5. Luc7 mutations affect splicing of *SUS1* pre-mRNA *in vivo*. RNA isolated from the indicated *LUC7* strains was reverse transcribed with an oligo(dT) primer. cDNAs were PCR-amplified with gene-specific sense and antisense primer pairs derived from the first and last exons of the *GLC7* (panel A) and *SUS1* (panel B) genes. The exon-intron organization of the pre-mRNAs is shown, with exons depicted as horizontal cylinders. The PCR products were resolved by native agarose gel electrophoresis and visualized by staining with ethidium bromide. The products of PCR amplification of genomic DNA with the same primer pairs are shown in lanes labeled "DNA". The positions and sizes (bp) of linear duplex DNA markers are indicated on the *left*. The RT-PCR products of unspliced, partly spliced, and fully spliced transcripts are specified on the *right*.

SUS1 in light of our prior findings that *SUS1* splicing is adversely affected by the Cbc2-Y24A mutation (Qiu et al., 2012) and by Msl5 mutations that perturb the Msl5•branchpoint RNA interface (Jacewicz et al., 2015). *SUS1* is one of the few yeast genes that contain two introns and it is the splicing of the first intron (which has a nonconsensus 5′ splice site GUAUG**A** and a nonconsensus branchpoint sequence UACU**G**AC) that is selectively impaired in the absence of CBC (Hossain et al., 2009).

cDNAs synthesized by oligo(dT)-primed reverse transcription of total RNA from yeast strains with wild-type and mutant *LUC7* alleles were PCR-amplified using gene specific primer pairs flanking the first and last introns, such that agarose gel electrophoresis would resolve the longer PCR fragments derived from unspliced pre-mRNA (which comigrates with the control fragment amplified from a genomic DNA template) and the shorter PCR fragments derived from mature spliced mRNA or, in the case of *SUS1*, partially spliced intermediates. None of the Luc7 mutations affected *GLC7* splicing (Fig. 2.5A).

As reported previously (Hossain et al., 2009; Jacewicz et al., 2015; Qiu et al., 2012), the *SUS1* transcripts in wild-type cells consisted predominantly of doubly spliced mature mRNA plus a minor component of partly spliced intermediate species (Fig. 2.5B). In *LUC7-(19–261)*, *-(31–261)*, and *R215A* cells, the mature mRNA was diminished, and the singly spliced intermediate comprised the majority species (Fig. 2.5B). These three *LUC7* alleles that exerted the greatest impact on *SUS1* splicing efficiency were also ones that exhibited the widest spectrum of synthetic lethalities and that were most effective in bypassing Prp28. *LUC7* alleles *C68A*, *D214A*, *R216A*, and *D219A* also affected *SUS1* splicing efficiency, manifest by a shift in the transcript distribution toward the partly spliced intermediate, albeit not to the point that it was the majority species detected by RT-PCR (Fig. 2.5B). *C68A*, *D214A*, *R216A*, and

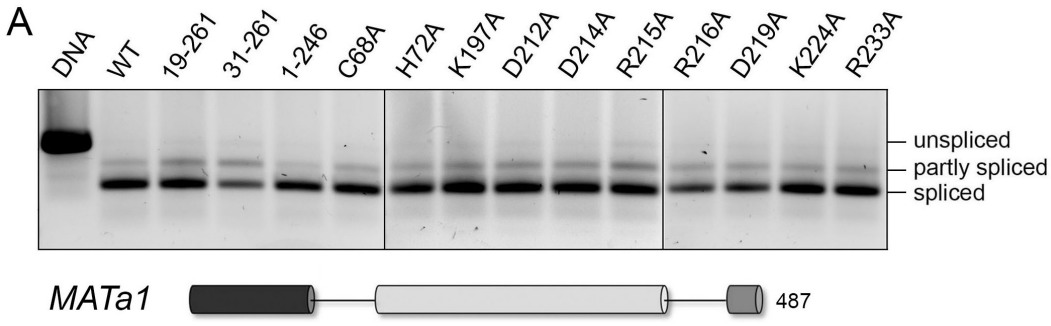
D219A all displayed multiple synthetic genetic interactions and were able to bypass Prp28. In contrast, the *LUC7-(1–246)*, *H72A*, *R233A*, and *K224A* alleles had little or no impact on *SUS1* splicing (as gauged by the relative amounts of fully spliced versus partly spliced species); these alleles were the ones that (i) displayed either no synthetic phenotypes—in the case of *LUC7-(1–246)*, *H72A*, and *R233A*—or narrowly mild mutation synergies (*K224A*); and (ii) were unable to bypass Prp28.

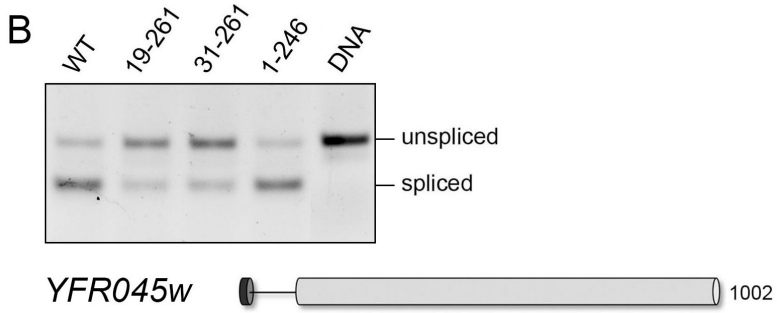
I surveyed the same wild-type and mutant *LUC7* strains for splicing of the *MATa1* pre-mRNA, which contains two introns, albeit both with consensus splicing signals. PCR amplification of the *MATa1* cDNA with primers in exon 1 and exon 3 revealed efficient removal of both introns in wildtype cells, where the mature spliced mRNA was the predominant species, with a minority fraction of partially spliced intermediate, and hardly any unspliced precursor (Fig. 2.6A). Whereas the *LUC7-(31–261)* strain had a diminished level of mature *MATa1* transcript, the other *LUC7* mutants maintain the wild-type distribution of mature and partially spliced species.

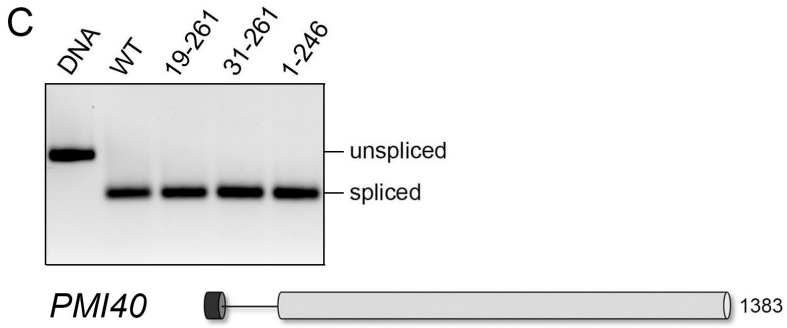
Finally, I evaluated the effects of the Luc7 N-terminal truncations on the splicing of the single-intron *YFR045w* and *PMI40* pre-mRNAs. *YFR045w*, which has a nonconsensus 5′ splice site GUA**A**GU and a nonconsensus branchpoint sequence UA**U**UAAC, was relatively inefficiently spliced in wild-type cells (e.g., compared to *GLC7*); unspliced *YFR045w* precursor was evident by RT-PCR, but was less abundant than the spliced mature transcript (Fig. 2.6B). I found that *YFR045w* splicing was impeded in *LUC7-(19-261)* and *LUC7-(31-261)* cells, where the mature *YFR045w* mRNA was diminished and the unspliced premRNA was now the majority species (Fig. 2.6B). (Note that *YFR045w* splicing in *LUC7-[1–246]* cells was no different from the wild-type strain [Fig. 2.6B].) In contrast, the *PMI40* pre-mRNA, which has a

nonconsensus branchpoint sequence (AACUAAC) but a consensus 5′SS, was spliced very efficiently in wild-type *LUC7* cells and was unaffected by the *LUC7* N-terminal truncations (Fig. 2.6C).

Figure 2.6. Effect of Luc7 mutations on splicing of *MATa1, YFR045w*, and *PMI40* pre-mRNAs. RNA isolated from the indicated *LUC7* strains was reverse transcribed with an oligo(dT) primer. cDNAs were PCR-amplified with gene-specific sense and antisense primer pairs derived from the first and last exons of the *MATa1* (panel A), *YFR045w* (panel B) and *PMI40* (panel C) genes (Jacewicz et al. 2015). The exon-intron organization of the pre-mRNAs are shown, with exons depicted as horizontal cylinders. The PCR products were resolved by native agarose gel electrophoresis and visualized by staining with ethidium bromide. The products of PCR amplification of genomic DNA with the same primer pairs are shown in lanes labeled "DNA". The RT-PCR products of unspliced, partly spliced, and fully spliced transcripts are specified on the *right*.


A
DNA
WT
19-261
31-261
1-246
C68A
H72A
K197A
D212A
D214A
R215A
R216A
D219A
K224A
R233A
unspliced
partly spliced
spliced
MATa1
487


B
WT
19-261
31-261
1-246
DNA
unspliced
spliced
YFR045w
1002


C
DNA
WT
19-261
31-261
1-246
unspliced
spliced
PMI40
1383

DISCUSSION

The present study affords new genetic insights into Luc7 function and structure–activity relations. I defined Luc7- (31–246) as a minimal functional unit for complementation of vegetative growth of a *luc7Δ* null strain. Though dispensable for growth in an otherwise wild-type background, the N-terminal peptide of Luc7 (^{1}MSTMSTPAAEQRKLVEQL^{18}) plays an important role in U1 snRNP function, which was evinced by our findings that the *LUC7-(19-261)* allele (i) impaired the splicing of *SUS1* and *YFR045w* pre-mRNAs in vivo; (ii) was synthetically lethal absent other U1 snRNP constituents (Mud1, Nam8, the TMG cap, the C terminus of Snp1), absent Mud2, and when the cap-binding site of nuclear CBC was debilitated; and (iii) bypassed the need for the essential DEAD-box ATPase Prp28. This constellation of effects implicates the N-terminal peptide in fulfilling two aspects of Luc7 function: in stability of the U1•pre-mRNA complex and in aiding the splicing of yeast pre-mRNAs that have a nonconsensus 5′SS element. The N-terminal peptide includes eight positions of side chain conservation in all four Luc7 proteins aligned in Figure 2.1.

Alanine scanning revealed that whereas mutations of the CCHH ZnF2 motif are lethal, mutations of the ZnF1 CCCH motif are not, signifying that the ZnF1 and ZnF2 motifs are not functionally redundant. Although the ZnF2 module bracketed by the essential putative metal-binding cysteines and histidines is conserved among the Luc7 homologs, none of the eight alanine changes introduced in lieu of conserved basic and acidic residues affected yeast growth. Yet, the alanine scan highlighted the functional importance of individual ZnF2 module constituents Asp214, Arg215, Arg216, and Asp219, as gauged by their effects on *SUS1* splicing, synergies with other splicing mutations, and Prp28 bypass suppression. The results suggest a scenario

in which a putative metal-coordination complex formed by the CCHH ZnF2 motif establishes a specific tertiary structure of the ^{201}CEVCGAYLSRLDTDRRLADHFLGKIH^{226} segment that is required for Luc7 activity, likely via contacts made by the conserved charged amino acids that our mutagenesis identified as relevant. It is noteworthy that Ester and Uetz (2008) had previously identified a 15-aa peptide within the ZnF2 module—^{214}DRRLADHFLGKIHLG^{228}—in their screen of an overlapping Luc7 peptide library for interaction with the tandem FF1-2 domains of yeast Prp40. Focusing further on binding of this peptide to the Prp40 FF1 domain, they conducted an alanine scan of every position in the Luc7 peptide. Most pertinent to our results, they reported that alanine substitutions for Asp214, Arg215, Arg216, Asp219, and His220 did not diminish Luc7 peptide interaction with the Prp40 FF1 domain (Ester and Uetz, 2008). I infer that the in vivo phenotypes I see for the *D214A*, *R215A*, *R216A*, *D219A*, and *H220A* alleles do not reflect a simple defect in the imputed Luc7•Prp40-FF1 interaction. Because Ester and Uetz (2008) did note that a *K224A* mutation of the Luc7 peptide nearly effaced its interaction with Prp40-FF1, it is conceivable that the mild in vivo mutational synergies noted here for the *LUC7-K224A* allele might arise via an effect on Luc7•Prp40-FF1 interaction.

Three of the candidate metal-ligands of the ZnF1 motif (Cys45, Cys53, and Cys68), though not individually essential for yeast survival, are functionally relevant by the following criteria: inefficient *SUS1* splicing (in *C68A* cells); temperature-sensitive growth defects (*C45A* and *C53A*); and severe synthetic phenotypes and Prp28 bypass (*C45A*, *C53A*, and *C68A*). His72 was unique in that its mutation to alanine elicited no effects in any of the tests I applied. Assuming that the ZnF1 motif really does coordinate a metal, it is possible that the place of the His72 ligand in the

metal coordination complex can be taken either by another (albeit nonconserved) histidine in the vicinity of the ZnF1 motif (e.g., His40 or His76) or by a water molecule.

The genetic analysis here establishes a foundation for understanding Luc7 function that will ultimately hinge on obtaining a crystal structure of yeast Luc7 per se and/or a cryo-EM structure of the complete yeast U1 snRNP.

ACKNOWLEDGEMENTS

Beate Schwer carried out the genetic interaction experiments for *snp1(1-233)*.

REFERENCES

Abovich, N., Liao, X.C., Rosbash, M., 1994. The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. Genes Dev 8, 843-854.

Abovich, N., Rosbash, M., 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell 89, 403-412.

Chang, J., Schwer, B., Shuman, S., 2012. Structure-function analysis and genetic interactions of the yeast branchpoint binding protein Msl5. Nucleic Acids Res 40, 4539-4552.

Chen, J.Y., Stands, L., Staley, J.P., Jackups, R.R., Jr., Latus, L.J., Chang, T.H., 2001. Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Mol Cell 7, 227-232.

Colot, H.V., Stutz, F., Rosbash, M., 1996. The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes Dev 10, 1699-1708.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., Prinz, J., St Onge, R.P., VanderSluis, B., Makhnevych, T., Vizeacoumar, F.J., Alizadeh, S., Bahr, S., Brost, R.L., Chen, Y., Cokol, M., Deshpande, R., Li, Z., Lin, Z.Y., Liang, W., Marback, M., Paw, J., San Luis, B.J., Shuteriqi, E., Tong, A.H., van Dyk, N., Wallace, I.M., Whitney, J.A., Weirauch, M.T., Zhong, G., Zhu, H., Houry, W.A., Brudno, M., Ragibizadeh, S., Papp, B., Pal, C., Roth, F.P., Giaever, G., Nislow, C., Troyanskaya, O.G., Bussey, H., Bader, G.D., Gingras, A.C., Morris, Q.D., Kim, P.M., Kaiser, C.A., Myers, C.L., Andrews, B.J., Boone, C., 2010. The genetic landscape of a cell. Science 327, 425-431.

Ester, C., Uetz, P., 2008. The FF domains of yeast U1 snRNP protein Prp40 mediate interactions with Luc7 and Snu71. BMC Biochem 9, 29.

Fortes, P., Bilbao-Cortes, D., Fornerod, M., Rigaut, G., Raymond, W., Seraphin, B., Mattaj, I.W., 1999a. Luc7p, a novel yeast U1 snRNP protein with a role in 5' splice site recognition. Genes Dev 13, 2425-2438.

Fortes, P., Kufel, J., Fornerod, M., Polycarpou-Schwarz, M., Lafontaine, D., Tollervey, D., Mattaj, I.W., 1999b. Genetic and physical interactions involving the yeast nuclear cap-binding complex. Mol Cell Biol 19, 6543-6553.

Gottschalk, A., Tang, J., Puig, O., Salgado, J., Neubauer, G., Colot, H.V., Mann, M., Seraphin, B., Rosbash, M., Luhrmann, R., Fabrizio, P., 1998. A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. RNA 4, 374-393.

Hage, R., Tung, L., Du, H., Stands, L., Rosbash, M., Chang, T.H., 2009. A targeted bypass screen identifies Ynl187p, Prp42p, Snu71p, and Cbp80p for stable U1 snRNP/Pre-mRNA interaction. Mol Cell Biol 29, 3941-3952.

Hausmann, S., Zheng, S., Costanzo, M., Brost, R.L., Garcin, D., Boone, C., Shuman, S., Schwer, B., 2008. Genetic and biochemical analysis of yeast and human cap trimethylguanosine synthase: functional overlap of 2,2,7-trimethylguanosine caps, small nuclear ribonucleoprotein components, pre-mRNA splicing factors, and RNA decay pathways. J Biol Chem 283, 31706-31718.

Hossain, M.A., Claggett, J.M., Nguyen, T., Johnson, T.L., 2009. The cap binding complex influences H2B ubiquitination by facilitating splicing of the SUS1 pre-mRNA. RNA 15, 1515-1527.

Jacewicz, A., Chico, L., Smith, P., Schwer, B., Shuman, S., 2015. Structural basis for recognition of intron branchpoint RNA by yeast Msl5 and selective effects of interfacial mutations on splicing of yeast pre-mRNAs. RNA 21, 401-414.

Jacewicz, A., Schwer, B., Smith, P., Shuman, S., 2014. Crystal structure, mutational analysis and RNA-dependent ATPase activity of the yeast DEAD-box pre-mRNA splicing factor Prp28. Nucleic Acids Res 42, 12885-12898.

Kao, H.Y., Siliciano, P.G., 1992. The yeast homolog of the U1 snRNP protein 70K is encoded by the SNP1 gene. Nucleic Acids Res 20, 4009-4013.

Kondo, Y., Oubridge, C., van Roon, A.M., Nagai, K., 2015. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. Elife 4.

Liao, X.L.C., Tang, J., Rosbash, M., 1993. An Enhancer Screen Identifies a Gene That Encodes the Yeast U1 Snrnp a-Protein - Implications for Snrnp Protein Function in Premessenger Rna Splicing. Gene Dev 7, 419-428.

Puig, O., Bragado-Nilsson, E., Koski, T., Seraphin, B., 2007. The U1 snRNP-associated factor Luc7p affects 5 ' splice site selection in yeast and human. Nucleic Acids Research 35, 5874-5885.

Qiu, Z.R., Chico, L., Chang, J., Shuman, S., Schwer, B., 2012. Genetic interactions of hypomorphic mutations in the m7G cap-binding pocket of yeast nuclear cap binding complex: an essential role for Cbc2 in meiosis via splicing of MER3 pre-mRNA. RNA 18, 1996-2011.

Qiu, Z.R., Schwer, B., Shuman, S., 2015. Two Routes to Genetic Suppression of RNA Trimethylguanosine Cap Deficiency via C-Terminal Truncation of U1 snRNP Subunit Snp1 or Overexpression of RNA Polymerase Subunit Rpo26. G3 (Bethesda) 5, 1361-1370.

Schwer, B., Chang, J., Shuman, S., 2013. Structure-function analysis of the 5' end of yeast U1 snRNA highlights genetic interactions with the Msl5*Mud2 branchpoint-binding complex and other spliceosome assembly factors. Nucleic Acids Res 41, 7485-7500.

Schwer, B., Erdjument-Bromage, H., Shuman, S., 2011. Composition of yeast snRNPs and snoRNPs in the absence of trimethylguanosine caps reveals nuclear cap binding protein as a gained U1 component implicated in the cold-sensitivity of tgs1Delta cells. Nucleic Acids Res 39, 6715-6728.

Schwer, B., Shuman, S., 2014. Structure-function analysis of the Yhc1 subunit of yeast U1 snRNP and genetic interactions of Yhc1 with Mud2, Nam8, Mud1, Tgs1, U1 snRNA, SmD3 and Prp28. Nucleic Acids Res 42, 4697-4711.

Schwer, B., Shuman, S., 2015. Structure-function analysis and genetic interactions of the Yhc1, SmD3, SmB, and Snp1 subunits of yeast U1 snRNP and genetic interactions of SmD3 with U2 snRNP subunit Lea1. RNA 21, 1173-1186.

Staley, J.P., Guthrie, C., 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Mol Cell 3, 55-64.

Wilmes, G.M., Bergkessel, M., Bandyopadhyay, S., Shales, M., Braberg, H., Cagney, G., Collins, S.R., Whitworth, G.B., Kress, T.L., Weissman, J.S., Ideker, T., Guthrie, C., Krogan, N.J., 2008. A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. Mol Cell 32, 735-746.

Chapter 3

Domain requirements and genetic interactions of the Mud1 subunit of the *Saccharomyces cerevisiae* U1 snRNP

INTRODUCTION

pre-mRNA splicing initiates when the U1 snRNP engages the intron 5' splice site (5'SS). The *Saccharomyces cerevisiae* U1 snRNP consists of a trimethylguanosine (TMG) capped 568-nt U1 snRNA, a seven-subunit Sm protein ring (present also in the U2, U4, and U5 snRNPs), and ten U1-specific protein subunits: Prp39, Prp40, Snu71, Snu56, Snp1, Mud1, Luc7, Prp42, Nam8, and Yhc1 (Fortes et al., 1999; Gottschalk et al., 1998; Schwer et al., 2011). Base-pairing of the U1 snRNA leader motif 5'-ACUUAC sequence with the consensus yeast 5'SS element 5'-GUAUGU nucleates an initial U1•pre-mRNA complex. Cross-intron bridging interactions between the yeast U1 snRNP at the 5'SS and the Msl5•Mud2 heterodimer at the branchpoint sequence 5'-UACUAAC then stabilize a commitment complex, which provides a scaffold for recruitment of the U2 snRNP to the branchpoint (Abovich and Rosbash, 1997).

Traditional genetics and synthetic genetic arraying, as well as structure-guided mutagenesis, have identified a rich network of genetically buffered functions during early spliceosome assembly in budding yeast, embracing the U1-specific snRNP proteins Mud1, Nam8, Yhc1, Snp1, and Luc7, the U1 snRNA, the TMG cap, the Cbc2•Sto1 nuclear m^7G cap-binding complex (CBC), the DEAD-box ATPase Prp28, the Msl5•Mud2 branchpoint-binding complex, the Msl1 and Lea1 subunit of the U2 snRNP, and the seven subunits of the Sm protein ring (Abovich et al., 1994; Agarwal et al., 2016; Chang et al., 2012; Colot et al., 1996; Costanzo et al., 2010; Gottschalk et al., 1998; Hausmann et al., 2008; Jacewicz et al., 2015; Liao et al., 1993; Qiu et al.,

2012; Schwer et al., 2013; Schwer et al., 2016; Schwer et al., 2017; Schwer and Shuman, 2014, 2015; Wilmes et al., 2008). This network is defined by the numerous instances in which null alleles of inessential players (e.g., Mud1, Nam8, Mud2) or benign mutations in essential factors (e.g., Yhc1, Snp1, Luc7 Msl5, Sm ring subunits) elicit synthetic lethal phenotypes when combined with other benign mutations in the splicing machinery.

The composition of the U1 snRNP is more complex in budding yeast than in mammals, with respect to the size of the U1 snRNA (568 *versus* 164 nt) and the number of U1-specific protein subunits (ten *versus* three) that are assimilated into U1 snRNP along with the 7-subunit Sm protein ring. Yeast Snp1, Mud1, and Yhc1, are the homologs of human U1-70K, U1-A and U1-C, respectively. Whereas Snp1 and Yhc1 are essential for yeast viability, Mud1 is not (Liao et al., 1993; Smith and Barrell, 1991; Tang et al., 1997). This contrasts with the essentiality of Mud1 homolog U1A for mammalian U1 snRNP function. U1A and Mud1 consist of two RRM (RNA recognition motif) domains separated by a linker peptide. U1A has been the subject of intense study, as a paradigm of RNA recognition by RRM-containing proteins (Maris et al., 2005). Indeed, the landmark crystal structure of the N-terminal RRM1 domain of U1A bound to the stem-loop II segment in human U1 snRNA (Oubridge et al., 1994) established key principles that apply broadly to the RRM family. U1A RRM1 specifically recognizes the sequence AUUGCAC atop stem-loop II in the mammalian U1 RNA. By contrast, the U1A C-terminal RRM2 domain does not bind RNA (Lu and Hall, 1997).

Mud1 does not recapitulate U1A in the respective yeast and human U1 snRNPs, insofar as the yeast U1 snRNA does not have an equivalent of the human U1 snRNA stem-loop II element to which U1A binds. Tang and Rosbash (1996)

employed *in vivo* dimethylsulfate (DMS) modification to probe the U1 snRNA in *MUD1* versus *mud1*Δ cells and thereby presented evidence for enhanced DMS accessibility to two regions of yeast U1 snRNA (nt 61 to 68 and 133 to 150, imputed to comprise opposing looped out segments within a duplex stem) when Mud1 is absent. As neither of these regions has any sequence similarity to the AUUGCAC site recognized by U1A, one of the following scenarios can be surmised: (i) if Mud1 binds yeast U1 snRNA, it does so with different specificity than U1A; (ii) Mud1 does not bind U1 snRNA directly and the DMS accessibility absent Mud1 reflects indirect effects on U1 snRNA conformation or on protein interaction within the U1 snRNP. Whereas Tang and Rosbash (1996) showed that deleting RRM2 of Mud1 had no effect on DMS sensitivity (implying that RRM2 is not essential for association of Mud1 with U1 snRNP), the ΔRRM2 version of Mud1 was biologically inactive in an *in vivo* reporter assay employing an in efficiently spliced synthetic intron that depends on Mud1. They proposed that RRM1 serves to tether RRM2 to the U1 snRNP.

Here I revisit the issue of how the domains of Mud1 are organized functionally. I exploit an expanded array of synthetic genetic interactions of Mud1 uncovered in the two decades since Mud1 was last studied. By surveying for complementation of synthetic lethality by Mud1 domains, *per se* and in various combinations, I show that RRM1 and RRM2 are both necessary for full Mud1 biological activity, though they need not be linked within the same polypeptide. Specifically, I find that co-expression of an RRM1-containing N-terminal fragment Mud1-(1-127) plus a C-terminal fragment Mud1-(128-298), comprising an interdomain linker and RRM2, was able to complement *mud1*Δ synthetic lethality in multiple genetic backgrounds. The linker segment, which I show includes a nuclear localization signal (NLS), was necessary to confer activity on RRM2. Guided by a

recent cryo-EM structure of yeast U1 snRNP (Li et al., 2017), I conducted an alanine scan of RRM1 amino acids that contact the U1 snRNA and found that none were essential *per se* for Mud1 activity *in vivo*.

EXPERIMENTAL PROCEDURES

MUD1 expression plasmids and mutants

Yeast plasmids pRS415-MUD1 (*CEN LEU2*) and pRS413-MUD1 (*CEN HIS3*) contain a *MUD1* expression cassette composed of the *MUD1* cDNA, 569-bp of 5'-flanking genomic DNA (containing the putative MUD1 promoter), and 309-bp of 3'-flanking genomic DNA. Restriction sites for SacI and XmaI were introduced at the 5' and 3' ends of the cassette to allow insertion between the SacI and XmaI sites of the pRS415 and pRS413 plasmids. A BamHI site was introduced immediately 5' of the *MUD1* translation start codon and a SpeI site was placed immediately 3' of the stop codon; these sites facilitated cloning of *MUD1* mutant alleles. N-terminal truncation variants were generated by PCR using forward primers that introduced a BamHI site followed by a Met residue in lieu of Lys127, Ile137, Lys147, Ala157, Lys177, and Glu196. C-terminal truncation variants were generated by PCR using reverse primers that introduced stop codons and a flanking SpeI site in lieu of Gly128 and Asn197. Alanine mutations were introduced into the full-length *MUD1* expression cassette by two-stage PCR overlap extension with mutagenic primers.

Plasmids for co-expression of Mud1 domain fragments were constructed as follows. First, the expression cassettes for N-terminal truncations *MUD1-(128-298)* and *MUD1-(197-298)* were PCR amplified from the pRS413 plasmids using a forward primer that introduced an XmaI site 569-bp upstream of the start codon (to replace the SacI site) and a reverse primer that introduced an ApaI site 309-bp downstream of the

stop codon (to replace the XmaI site). The PCR products were digested with XmaI and ApaI and inserted between XmaI and ApaI sites of p415-MUD1-(1-127) and p415-MUD1-(1-196) to generate plasmids p415-MUD1-(1-127+128-298), p415-MUD1-(1-196+197-298), and p415-MUD1-(1-127+197-298) with the two *MUD1* alleles arranged in a head-to-tail configuration. The *MUD1* genes in the expression plasmids were sequenced completely to confirm that no unwanted changes were acquired during amplification and cloning.

Tests of Mud1 function in vivo

I employed plasmid shuffle assays to test the effects of Mud1 mutations on complementation of *mud1*Δ synthetic lethality with *mud2*Δ, *nam8*Δ, *MSL5-L169A*, *MSL5-I189A*, and *MSL5-Y100A* (Chang et al., 2010; Jacewicz et al., 2015; Qiu et al., 2011). The yeast strains used for plasmid shuffle are listed in Table 1. To establish the complementation assay in the *msl1*Δ background, we prepared a heterozygous *MSL1 msl1::natMX MUD1 mud1::kanMX* diploid by crossing a *msl1::natMX* haploid (generated by integration of a *natMX* cassette flanked by 500 bp and 464 bp segments of genomic DNA upstream and downstream of the *MSL1* ORF) with *mud1::kanMX* cells of the opposite mating type. The diploids were then transformed with plasmid p360-MUD1 [*CEN URA3 MUD1*]. Ura^+ diploids were selected and sporulated. Asci were dissected and the desired haploid *mud1*Δ *msl1*Δ progeny were Ura^+ and resistant to clonNAT and geneticin. When *mud1*Δ *msl1*Δ p360-MUD1 cells were transferred to agar medium containing 5-fluoroorotic acid (FOA) to select against the *URA3 MUD1* plasmid, only tiny colonies were recovered after 6 days of incubation at 30°C. The resulting *mud1*Δ s *msl1*Δ strain failed to form colonies on YPD agar (Fig. 3.1).

To assay bioactivity of wild-type and mutated *MUD1* alleles, the double-mutant strains were transfected with *CEN LEU2 MUD1* plasmids. In case of *mud1*Δ *msl5*Δ,

MUD1 alleles on *CEN LEU2* plasmids were co-transfected with *MSL5-Ala* alleles on *CEN HIS3* plasmids. Transformants were selected and streaked on agar medium containing FOA. The plates were incubated at 20, 30, and 37°C, and alleles that failed to give rise to macroscopic colonies at any temperature after 8 days were deemed lethal. Individual FOA-resistant colonies with viable *MUD1* alleles were grown to mid-log phase in YPD broth and adjusted to A_{600} of 0.1. Aliquots (3 µl) of serial 10-fold dilutions were spotted on YPD agar plates, which were then incubated at temperatures ranging from 20 to 37˚C.

Table 3.1. Yeast strains used in this study

Strain	Genotype	Reference/Source
yRJX7	*MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 mud2::natMX mud1::kanMX* p360-MUD1 [*CEN URA3 MUD1*]	(Chang et al., 2012)
yRJBB42	*MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 nam8Δ::natR mud1Δ::kanMX* p360-NAM8 [*CEN URA3 NAM8*]	(Qiu et al., 2011)
yRJMM23	*MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 msl1::natMX mud1::kanMX* p360-MUD1 [*CEN URA3 MUD1*]	This study
yRJGG8	*MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 msl5::natMX mud1::kanMX* p360-MSL5 [*CEN URA3 MSL5*]	(Schwer et al., 2013)
W303a	*MATa leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 [phi$^+$]*	(Rothstein, 1983)

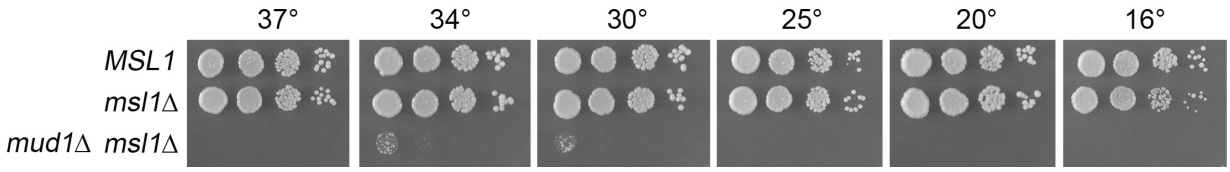


Figure 3.1. *mud1*Δ is synthetically lethal with *msl1*Δ. Exponentially growing wild-type *MSL1* and *msl1*Δ cells were tested for growth in parallel with *mud1*Δ *msl1*Δ cells recovered after FOA selection and transfer to liquid medium. The cultures were adjusted to A_{600} of 0.1 and aliquots (3 µl) of serial 10-fold dilutions were spotted on YPD agar plates. The plates were photographed after incubation at the indicated temperatures.

Mud1 linker-GFP reporters of cellular localization

Two-stage PCR overlap extension was used to create DNA fragments in which the Mud1 linker segments aa 128-196, aa 128-154, and aa 155-196 were fused in frame to the N-terminus of green fluorescent protein (GFP). pYN132-GFP was used as the template to amplify the GFP open reading frame (Schneider and Schwer, 2001). In the process, restriction sites were introduced for cloning of the DNA segments into a pRS423-based plasmid in which expression of GFP and the linker-GFP fusions is under the transcriptional control of the TPI1 promoter. The *2μ HIS3* plasmids were transfected into *W303a* cells, transformants were selected and grown in SD-His media to A_{600} of 0.6. Aliquots of the cultures were harvested by centrifugation and resuspended in 4% paraformaldehyde for fixation. After washing with phosphate buffered saline (PBS), the cells were stained with DAPI. The cells were then washed again with PBS and mounted to a glass slide. Images were captured using an Axio2 microscope (Zeiss) using a 100x objective connected to a charge-coupled device camera and processed using SlideBook software (Intelligent Imaging Innovations).

RESULTS

Complementation of synthetic lethal mud1Δ phenotypes by expression of Mud1 domains

Yeast Mud1 and human U1A share an RRM1-linker-RRM2 domain organization but are not highly similar at the amino acid sequence level. By contrast, a primary structure alignment of the 298-aa *S. cerevisiae* Mud1 protein to the homologous 296-aa protein from the fungus *Vanderwaltozyma polyspora* highlights 167 positions of side chain identity/similarity (Fig. 3.2). I used this alignment as a guide to divide Mud1 into the following segments: RRM1 (aa 1-127); linker-RRM2

(aa 128-298); RRM1-linker (aa 1-196); and RRM2 (aa 197-298). *CEN* plasmids bearing wild-type *MUD1* or truncated *MUD1* alleles under the control of the native *MUD1* promoter were tested by plasmid shuffle for complementation of *mud1*Δ in six different double-mutant genetic backgrounds in which *mud1*Δ is synthetically lethal. These included: (i) absence of U1 snRNP subunit Nam8 (Qiu et al., 2011), branchpoint binding protein subunit Mud2 (Abovich et al., 1994; Chang et al., 2010), and U2 snRNP subunit Msl1; and (ii) alanine mutations of Msl5 within the intron branchpoint RNA-binding domain (L169A, I189A) or in the Mud2-interaction domain (Y100A) (Schwer et al., 2013). Expression of full-length wild-type Mud1 complemented the lethality of all test strains. Expression of RRM1 or RRM2 alone failed to provide Mud1 function in any genetic background tested (Fig. 3.3). Restoring the linker segment to RRM2 elicited a gain of partial function, whereby linker-RRM2 expression weakly complemented *mud1*Δ in the context of *MSL5-Y100A* (scored as ± and temperature-sensitive *ts*) (Fig. 3.3) and *nam8*Δ (+ growth and cold-sensitive; *cs*) (Fig. 3.3 and 3.4). Appending the linker to RRM1 allowed *cs* complementation of *nam8*Δ (Fig. 3.3 and 3.4).

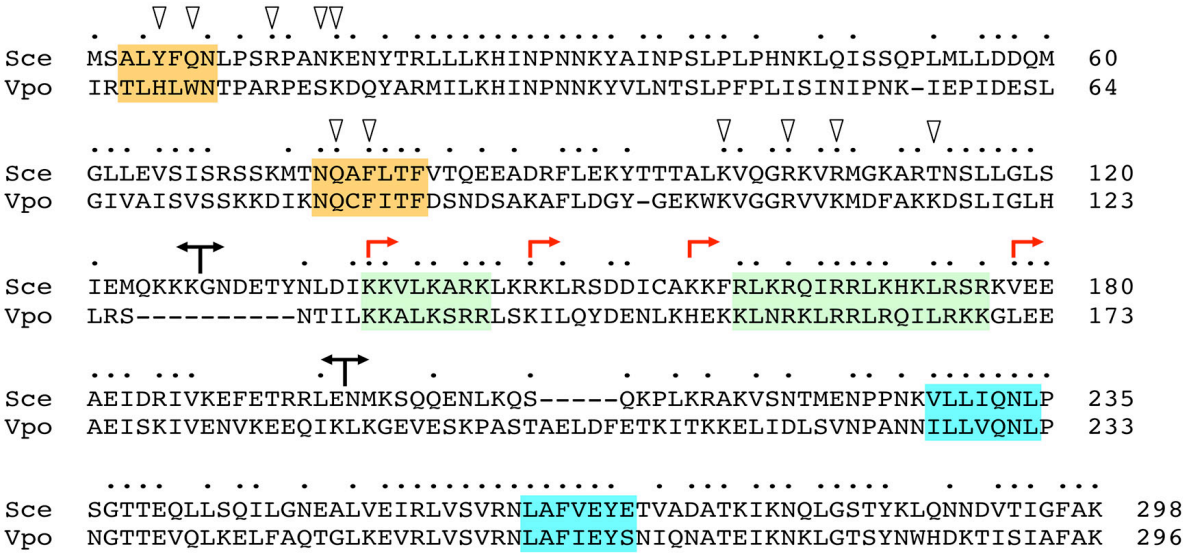


Figure 3.2. *Saccharomomyces* cerevisiae Mud1. The primary structures of the Mud1 polypeptides of *Saccharomyces cerevisiae* (Sce) and *Vanderwaltozyma polyspora* (Vpo) are aligned. Positions of side chain identity/similarity are indicated by •. Gaps in the alignment are denoted by dashes. Forward and reverse arrowheads indicate the boundaries of S. cerevisiae Mud1 N-terminal and C-terminal truncations, respectively, that were generated in the present study. The putative RNP motifs of the RRM1 and RRM2 domains are shaded gold and blue, respectively. Two blocks of basic amino acids in the linker are shaded green. The Mud1 amino acids subjected to alanine mutagenesis here are indicated by triangles.

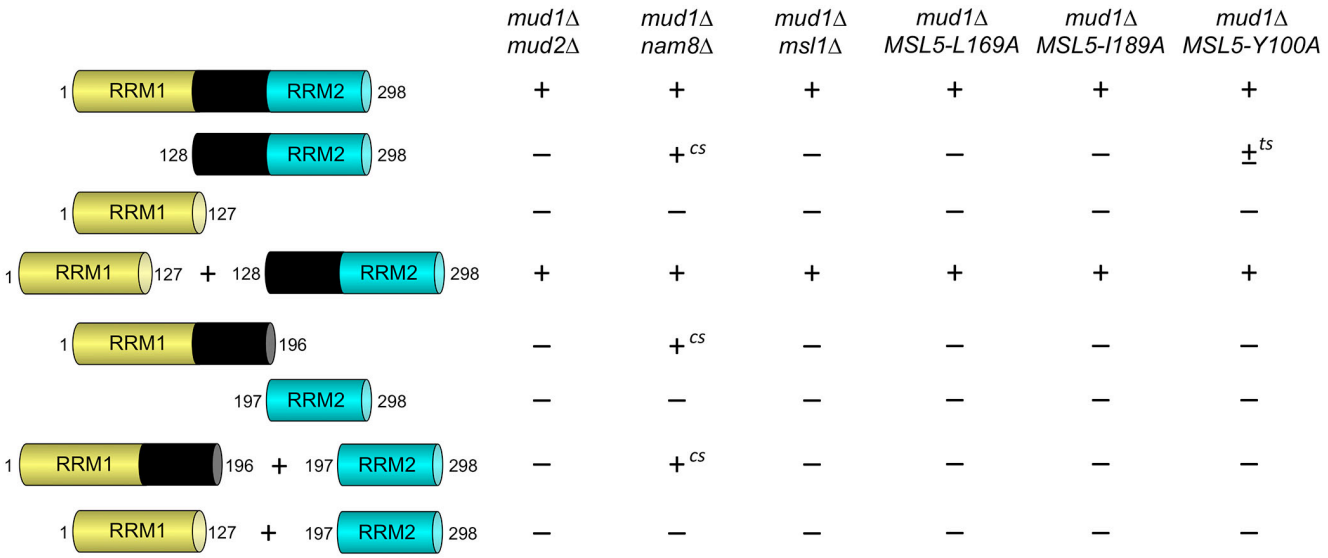


Figure 3.3. Mud1 domain requirements in multiple synthetic lethal genetic backgrounds. Wild-type and truncated *MUD1* alleles encoding the indicated Mud1 polypeptides were tested for activity by plasmid shuffle in six different double-mutant genetic backgrounds in which *MUD1* deletion is lethal. *MUD1* alleles, and combinations thereof, that failed to support growth of the test strains on FOA were deemed lethal (scored as – growth). The viable FOA-resistant strains expressing the indicated Mud1 proteins were spot-tested for growth on YPD agar at temperatures from 20˚C to 37˚C. Growth was scored as follows: (+) colony size indistinguishable from strains bearing wild-type *MUD1*; ($+^{cs}$) + growth at 37˚C and no growth at 20˚C; ($\pm^{ts}$) no growth at 37˚C and small colonies at other temperatures.

The salient finding was that co-expression of RRM1 and linker-RRM2 as separate proteins complemented Mud1 function in all six synthetic lethal genetic backgrounds (Fig. 3.3). Thus, whereas the RRM1 and RRM2 domains are both required for Mud1 activity, they need not be linked in *cis* within the same polypeptide. Additional insight was gained from the observations that: (i) co-expression of RRM1 and RRM2 (i.e. devoid of linker) failed to support growth of the six test strains; and (ii) co-expressing RRM1-linker and RRM2 was no better than expressing RRM1-linker alone (Fig. 3.3). These results indicate that the linker is needed in cis for activity of the RRM2 domain.

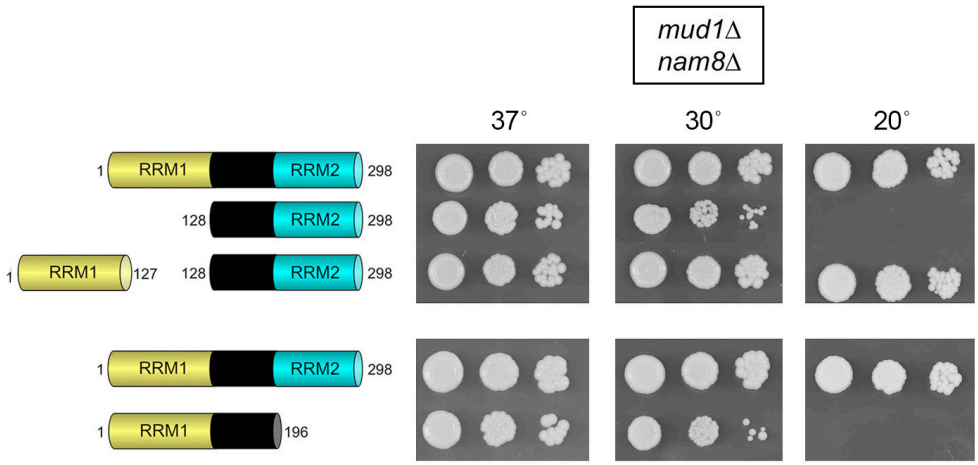


Figure 3.4. Complementation of *mud1*Δ *nam8*Δ. Wild-type and truncated *MUD1* alleles encoding the indicated Mud1 polypeptides were tested by plasmid shuffle for *mud1*Δ *nam8*Δ complementation. The viable FOA-resistant strains expressing the indicated Mud1 proteins were spot-tested for growth on YPD agar at the temperatures specified.

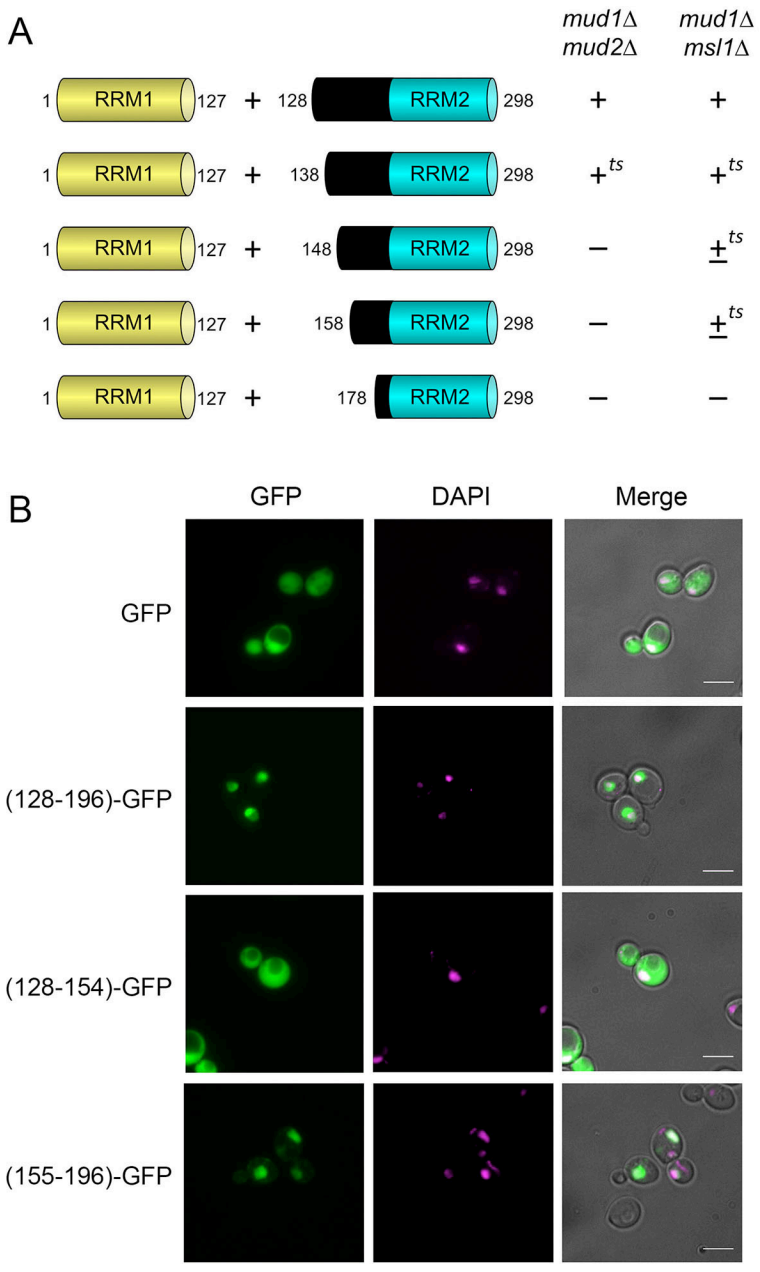
Probing the requirements for linker domain function

The Mud1 linker domain is rich in arginines and lysines (25/69 positions; Fig. 3.2), which are distributed in several clusters suggestive of a nuclear localization signal (NLS). Indeed, the linker domain of mammalian U1A is also arginine-lysine rich and has been shown to function as a long and complex NLS for U1A (Hetzer and Mattaj, 2000; Hieda et al., 2001; Kambach and Mattaj, 1992). To begin to delineate the important features of the linker domain I serially truncated the original linker-RRM2 protein (aa 128-298) to generate Mud1 constructs 138-298, 148-298, 158-298, and 178-298 (Fig. 3.2, N-termini indicated by red arrows). These alleles were co-expressed with RRM1 and tested for complementation in two genetic backgrounds in which the linker was found to be essential for RRM2 activity: *mud1*Δ *mud2*Δ and *mud1*Δ *msl1*Δ. Expression of the Mud1-(138-298) protein in tandem with RRM1 sustained growth of *mud1*Δ *mud2*Δ and *mud1*Δ *msl1*Δ cells at 20 to 30˚C but not at 37˚C (scored as $+^{ts}$ in Fig. 3.5A). The Mud1-(148-298), -(158-298), and -(178-298) constructs failed to complement *mud1*Δ *mud2*Δ when co-expressed with RRM1, signifying that the Mud1 segment from aa 138 to 147 is essential in this genetic background. By contrast, Mud1-(148-298) or Mud1-(158-298) plus RRM1 complemented the *mud1*Δ *msl1*Δ strain and conferred a slow growth at 20 to 30˚C and no growth at 37˚C (scored as $\pm^{ts}$ in Fig. 3.5A). This result underscores that the linker requirements vary in different genetic contexts. A further deletion of the linker segment from aa 158 to 177 abolished *mud1*Δ *msl1*Δ complementation (Fig. 3.5A).

To test if the linker can function as an NLS, I created a series of plasmid-borne Mud1 linker-GFP reporters in which the linker segments from aa 128-196, 128-154, or 155-196 were fused to the N-terminus of green fluorescent protein (GFP). These linker-GFP plasmids, and a control plasmid expressing GFP, were introduced into

wild-type yeast cells and the distribution of GFP was gauged by fluorescence microscopy. Whereas the GFP control was distributed throughout the yeast cell (exclusive of the vacuole), the (128-198)-GFP fusion was concentrated in the nucleus (Fig. 3.5B). Upon bisecting the linker in the fusion construct, I saw that whereas (128-154)-GFP was distributed diffusely (*à la* the GFP control), (155-196)-GFP localized to the nucleus (Fig. 3.5B). I conclude that an NLS resides within the Mud1 linker segment from aa 155-196, likely by virtue of the basic patch ^{161}RLKRQIRRLKHKLRSR^{176} that is conserved in *V. polyspora* Mud1 (Fig. 3.2). I infer that the necessity of this segment for Mud1 function in complementing *mud1*Δ synthetic lethality (Fig. 3.5A) reflects its capacity to direct the Mud1 RRM2 to the nucleus wherein splicing occurs.

Figure 3.5. Function of the Mud1 linker region. (A) The indicated linker-RRM2 proteins plus RRM1 were co-expressed and tested by plasmid shuffle for complementation of *mud1*Δ *mud2*Δ and *mud1*Δ *msl1*Δ strains. Combinations that failed to support growth of the test strains on FOA were deemed lethal (scored as – growth). The viable FOA-resistant strains expressing the indicated Mud1 proteins were spot-tested for growth on YPD agar at temperatures from 20˚C to 37˚C. Growth was scored as follows: (+) colony size indistinguishable from strains bearing wild-type *MUD1*; ($+^{ts}$) + growth at 20 and 25˚C and no growth at 37˚C; ($\pm^{ts}$) no growth at 37˚C and small colonies at other temperatures. (B) The linker acts as a nuclear localization signal (NLS). The linker region and its C- and N- terminal truncation*s* were fused to the N-terminus of GFP. These fusion proteins were expressed in *W303a* cells. Cells were observed by using differential interference contrast microscopy. GFP (green) was detected by fluorescence microscopy. The nuclei (magenta) was visualized with DAPI. The scale bar denotes 5 µm.


A
mud1Δ mud2Δ
mud1Δ msl1Δ
1 RRM1 127 + 128 RRM2 298
+
+
1 RRM1 127 + 138 RRM2 298
+ts
+ts
1 RRM1 127 + 148 RRM2 298
−
±ts
1 RRM1 127 + 158 RRM2 298
−
±ts
1 RRM1 127 + 178 RRM2 298
−
−
B
GFP
DAPI
Merge
GFP
(128-196)-GFP
(128-154)-GFP
(155-196)-GFP

Alanine scanning mutagenesis of Mud1 RRM1

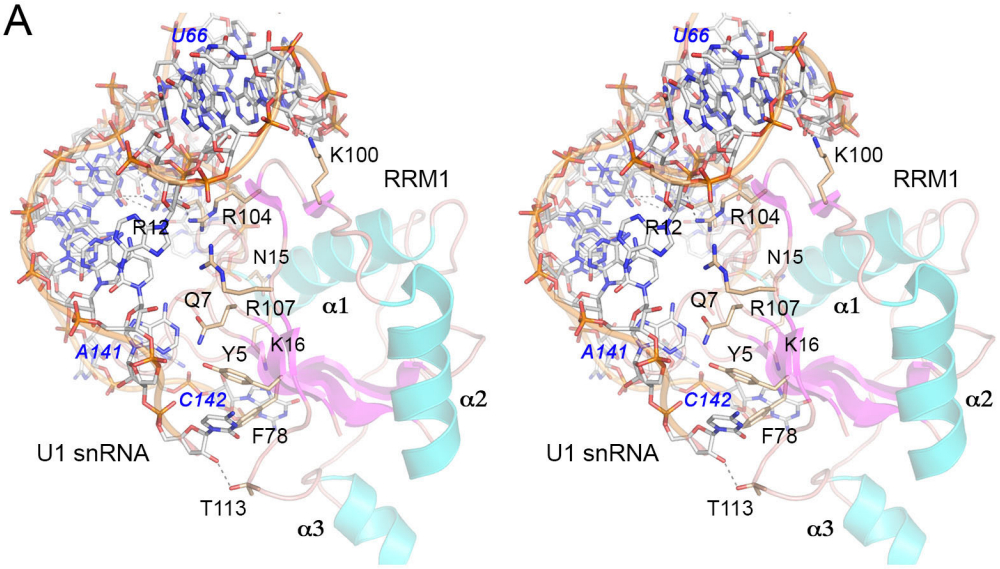
I undertook alanine scanning mutagenesis of *S. cerevisiae* Mud1 RRM1, guided initially by primary structure similarity to *V. polyspora* Mud1 (Fig. 3.2) and other fungal homologs and by presumptive similarities to the RRM1 component of U1A (Oubridge et al., 1994), the goal being to probe the effects of mutations at the putative Mud1•U1 snRNA interface. The subsequent report of the cryo-EM structure of the *S. cerevisiae* U1 snRNP allowed us to focus the findings I report here on mutations of Mud1 amino acids that interact with the U1 snRNA (Li et al., 2017)). The deposited U1 snRNP model (pdb 5UZ5) is missing large segments of many of the component proteins; in the case of Mud1, the model includes segments from aa 3 to 45, 55 to 125, and 132 to 148, i.e., there is no modeled structure of RRM2 or the essential portion of the linker (or the TAP-tag on the Mud1 C-terminus that was used by Li et al. to affinity-purify the U1 snRNP). The cryo-EM model of Mud1 RRM1 comprises a four-strand anti-parallel β sheet, a β hairpin loop, and three α helices (Fig. 4A). RRM1 contacts the U1 snRNA segments from nucleotides U^{57} to U^{66}, C^{140} to C^{142}, and G^{149} to G^{150} (Fig. 3.6A). It is especially noteworthy that these Mud1 contacts to U1 snRNA in the EM structure model correspond to the sites of enhanced DMS modification of the U1 snRNA in *mud1*Δ cells that were identified by Tang and Rosbash (1996) as the putative U1 snRNA-binding sites for Mud1 RRM1.

Here I introduced alanine substitutions at ten Mud1 amino acids that contact the U1 snRNA: Tyr5, Gln7, Arg12, Asn15, Lys16, Phe78, Lys100, Arg104, Arg107, and Thr113. Their contacts are indicated in Fig. 3.6B and illustrated in a stereo view of the RRM1-RNA interface in Fig. 3.6A. I also introduced alanine at Gln76 in the RNP1 motif of RRM1 (**Q**AFLTF; shaded gold in Fig. 3.2); this side chain is conserved in the RNP1 motif of *V. polyspora* Mud1 (**Q**CFITF) and human U1A (**Q**AFVIF). The

human U1A Gln side chain makes a π stacking interaction with purine nucleobase G9 of human U1 snRNA (Oubridge et al., 1994). The Mud1 Gln76 side chain is not modeled beyond the β-carbon in the yeast U1 snRNP structure. Single, double, and triple alanine mutations were introduced into the full-length *MUD1* gene and the mutant alleles were assayed by plasmid shuffle for *mud1*Δ *mud2*Δ complementation (Fig. 3.6B). All six of the single-alanine mutants – *Y5A*, *Q7A*, *R104A*, *T113A*, *Q76A*, and *F78A* – were as effective as wild-type *MUD1* in supporting growth of *mud1*Δ *mud2*Δ cells at 20 to 37˚C as was the *Y5A-Q7A* and *Q76A-F78A* double mutants (Fig. 3.6B). *mud1*Δ *mud2*Δ cells complemented by triple mutants *R12A-N15A-K16A* and *K100A-R014A-R107A* grew well at 20 to 34˚C but formed small colonies at 37˚C (scored as $+^{ts}$ in Fig. 3.6B).

Figure 3.6. Alanine scanning mutagenesis of Mud1. (A) Stereo view of the modeled Mud1 RRM1 domain and its interaction with U1 snRNA, prepared in Pymol using data from the cryo-EM structure in pdb 5UZ5. Mud1 is depicted as a cartoon trace with magenta β strands and cyan α helices. Amino acids subjected to alanine scanning are rendered as stick models with beige carbons. The segment of U1 snRNA with which Mud1 interacts is depicted as a stick model with gray carbons and a gold cartoon trace through the phosphodiester backbone. (B) Summary of the mutational analysis. Mud1 amino acids and their U1 snRNA contacts are listed in the first two columns on the left. Single, double, or triple alanine mutants of full-length Mud1 (listed in the third column) were tested by plasmid shuffle for *mud1*Δ *mud2*Δ complementation (fourth column at right).

A



B

Mud1 aa	U1 RNA contact	Mutation	*mud1*Δ *mud2*Δ complementation
Tyr5	A^{141} adenine, π-stack on C^{142} base	Y5A	+
Gln7	A^{141} adenine	Q7A Y5A-Q7A	+ +
Arg12	$^{60}CpA^{61}$ phosphate, π-cation on C^{60} base G^{134} guanine O6/N7	R12A-N15A-K16A	$+^{ts}$
Asn15	$^{57}UpU^{58}$ phosphate		
Lys16	π-cation on G^{149} base		
Gln76		Q76A	+
Phe78	C^{142} cytosine	F78A Q76A-F78A	+ +
Lys100	A^{67} ribose O2'		
Arg104	$^{62}ApU^{63}$ phosphate	R104A	+
Arg107	A^{64} adenine	K100A-R104A-R107A	$+^{ts}$
Thr113	C^{142} ribose O2'	T113A	+

DISCUSSION

The present study affords new genetic insights into Mud1 domain organization and structure-activity relations. By taking advantage of multiple genetic backgrounds in which Mud1 is essential for yeast vegetative growth, I delineated three Mud1 domains – RRM1, linker, and RRM2 – that are organized as two functional units: RRM1 and linker-RRM2, respectively. Tang and Rosbash (1997) were prescient in their deductions from DMS footprinting of the sites of contact between Mud1 and the U1 snRNA and their attribution of U1 snRNA contacts to the RRM1 domain, which apparently sufficed to be recruited to the U1 snRNP in the absence of RRM2. The key finding here is that RRM1 and linker-RRM2 need not be linked in *cis* within the same polypeptide to fulfill all Mud1 functions for vegetative growth, i.e., co-expression of RRM1 and linker-RRM2 as separate polypeptides complemented *mud1*Δ in every strain background. This belies the model suggested initially that RRM1 serves to tether the RRM2 to the U1 snRNP. Instead, our data indicate that the linker domain, which is needed in *cis* for activity of the RRM2 domain, directs RRM2 to the U1 snRNP independent of RRM1. The linker segment immediately flanking RRM2 contains an NLS that can guide the linker-RRM2 unit to the nucleus. The linker domain might also deliver RRM2 to the U1 snRNP, e.g., by virtue of interactions of the tracts of basic amino acids in the linker (Fig. 3.2) with U1 snRNA or other protein subunits of the U1 snRNP. The linker segment ^{132}TYNLDIKKVLKARKLKR^{148} that embraces the upstream basic tract is modeled in cryo-EM structures of the U1 snRNP and the *S. cerevisiae* pre-spliceosome as a long α helix (Bai et al., 2018; Li et al., 2017; Plaschka et al., 2018). However, this α helix is not in direct contact with the U1 snRNA and is remote from other protein subunits. The disposition of the linker NLS segment and RRM2 within the U1 snRNP is presently uncharted.

Differences in the ability of domain fragments of Mud1 to complement the growth of six synthetically lethal *mud1*Δ strains suggests a contextual hierarchy of Mud1 necessity ranging from: (i) most exigent, as in the cases of *mud1*Δ *mud2*Δ, *mud1*Δ *msl1*Δ, *mud1*Δ *MSL5-L169A*, and *mud1*Δ *MSL5-I189A* in which only the combination of RRM1 and linker-RRM2 sustains growth; to (ii) less stringent, exemplified by *mud1*Δ *nam8*Δ, in which either RRM1-linker or linker-RRM2 alone can restore viability, albeit with *cs* phenotypes (Fig. 3.3).

I conclude from the structure-guided alanine scan of the RRM1 domain that: (i) none of the individual Mud1 RRM1 amino acid contacts to U1 snRNA in the cryo-EM model is necessary for Mud1 activity in an exigent genetic background (*mud1*Δ *mud2*Δ); (ii) loss of both contacts of Tyr5 and Gln7 to the A141 adenine is benign; and (iii) triple mutations that eliminate multiple RNA contacts (e.g., to nucleotides U^{57} to A^{61} or A^{62} to A^{67}) are viable albeit *ts*.

The results here, and the recent structural studies of the yeast U1 snRNP, underscore significant differences between yeast Mud1 and human U1A, to the point that they might best be viewed as structural homologs (i.e., composed of two RRMs separated by a linker) but not as functional orthologs, notwithstanding a shared affiliation with their respective U1 snRNPs. The salient points are that Mud1 and U1A recognize entirely different sites in the yeast and human U1 snRNAs (Li et al., 2017; Oubridge et al., 1994) and their contributions to U1 snRNP function diverge genetically, with human U1A being essential and Mud1 being inessential.

ACKNOWLEDGEMENT

Beate Schwer generated the MUD1 cDNA plasmid and generated the *mud1*Δ *msl1*Δ strain.

REFERENCES

Abovich, N., Liao, X.C., Rosbash, M., 1994. The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. Genes Dev 8, 843-854.

Abovich, N., Rosbash, M., 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell 89, 403-412.

Agarwal, R., Schwer, B., Shuman, S., 2016. Structure-function analysis and genetic interactions of the Luc7 subunit of the Saccharomyces cerevisiae U1 snRNP. RNA 22, 1302-1310.

Bai, R., Wan, R., Yan, C., Lei, J., Shi, Y., 2018. Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation. Science 360, 1423-1429.

Chang, J., Schwer, B., Shuman, S., 2010. Mutational analyses of trimethylguanosine synthase (Tgs1) and Mud2: proteins implicated in pre-mRNA splicing. RNA 16, 1018-1031.

Chang, J., Schwer, B., Shuman, S., 2012. Structure-function analysis and genetic interactions of the yeast branchpoint binding protein Msl5. Nucleic Acids Res 40, 4539-4552.

Colot, H.V., Stutz, F., Rosbash, M., 1996. The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes Dev 10, 1699-1708.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., Prinz, J., St Onge, R.P., VanderSluis, B., Makhnevych, T., Vizeacoumar, F.J., Alizadeh, S., Bahr, S., Brost, R.L., Chen, Y., Cokol, M., Deshpande, R., Li, Z., Lin, Z.Y., Liang, W., Marback, M., Paw, J., San Luis, B.J., Shuteriqi, E., Tong, A.H., van Dyk, N., Wallace, I.M., Whitney, J.A., Weirauch, M.T., Zhong, G., Zhu, H., Houry, W.A., Brudno, M., Ragibizadeh, S., Papp, B., Pal, C., Roth, F.P., Giaever, G., Nislow, C., Troyanskaya, O.G., Bussey, H., Bader, G.D., Gingras, A.C., Morris, Q.D., Kim, P.M., Kaiser, C.A., Myers, C.L., Andrews, B.J., Boone, C., 2010. The genetic landscape of a cell. Science 327, 425-431.

Fortes, P., Bilbao-Cortes, D., Fornerod, M., Rigaut, G., Raymond, W., Seraphin, B., Mattaj, I.W., 1999. Luc7p, a novel yeast U1 snRNP protein with a role in 5' splice site recognition. Genes Dev 13, 2425-2438.

Gottschalk, A., Tang, J., Puig, O., Salgado, J., Neubauer, G., Colot, H.V., Mann, M., Seraphin, B., Rosbash, M., Luhrmann, R., Fabrizio, P., 1998. A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. RNA 4, 374-393.

Hausmann, S., Zheng, S., Costanzo, M., Brost, R.L., Garcin, D., Boone, C., Shuman, S., Schwer, B., 2008. Genetic and biochemical analysis of yeast and human cap trimethylguanosine synthase: functional overlap of 2,2,7-trimethylguanosine caps, small nuclear ribonucleoprotein components, pre-mRNA splicing factors, and RNA decay pathways. J Biol Chem 283, 31706-31718.

Hetzer, M., Mattaj, I.W., 2000. An ATP-dependent, Ran-independent mechanism for nuclear import of the U1A and U2B" spliceosome proteins. J Cell Biol 148, 293-303.

Hieda, M., Tachibana, T., Fukumoto, M., Yoneda, Y., 2001. Nuclear import of the U1A splicesome protein is mediated by importin alpha /beta and Ran in living mammalian cells. J Biol Chem 276, 16824-16832.

Jacewicz, A., Chico, L., Smith, P., Schwer, B., Shuman, S., 2015. Structural basis for recognition of intron branchpoint RNA by yeast Msl5 and selective effects of interfacial mutations on splicing of yeast pre-mRNAs. RNA 21, 401-414.

Kambach, C., Mattaj, I.W., 1992. Intracellular distribution of the U1A protein depends on active transport and nuclear binding to U1 snRNA. J Cell Biol 118, 11-21.

Li, X., Liu, S., Jiang, J., Zhang, L., Espinosa, S., Hill, R.C., Hansen, K.C., Zhou, Z.H., Zhao, R., 2017. CryoEM structure of Saccharomyces cerevisiae U1 snRNP offers insight into alternative splicing. Nat Commun 8, 1035.

Liao, X.C., Tang, J., Rosbash, M., 1993. An enhancer screen identifies a gene that encodes the yeast U1 snRNP A protein: implications for snRNP protein function in pre-mRNA splicing. Genes Dev 7, 419-428.

Lu, J., Hall, K.B., 1997. Tertiary structure of RBD2 and backbone dynamics of RBD1 and RBD2 of the human U1A protein determined by NMR spectroscopy. Biochemistry 36, 10393-10405.

Maris, C., Dominguez, C., Allain, F.H., 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 272, 2118-2131.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., Nagai, K., 1994. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature 372, 432-438.

Plaschka, C., Lin, P.C., Charenton, C., Nagai, K., 2018. Prespliceosome structure provides insights into spliceosome assembly and regulation. Nature 559, 419-422.

Qiu, Z.R., Chico, L., Chang, J., Shuman, S., Schwer, B., 2012. Genetic interactions of hypomorphic mutations in the m7G cap-binding pocket of yeast nuclear cap binding complex: an essential role for Cbc2 in meiosis via splicing of MER3 pre-mRNA. RNA 18, 1996-2011.

Qiu, Z.R., Schwer, B., Shuman, S., 2011. Determinants of Nam8-dependent splicing of meiotic pre-mRNAs. Nucleic Acids Res 39, 3427-3445.

Rothstein, R.J., 1983. One-step gene disruption in yeast. Methods Enzymol 101, 202-211.

Schneider, S., Schwer, B., 2001. Functional domains of the yeast splicing factor Prp22p. J Biol Chem 276, 21184-21191.

Schwer, B., Chang, J., Shuman, S., 2013. Structure-function analysis of the 5' end of yeast U1 snRNA highlights genetic interactions with the Msl5*Mud2 branchpoint-binding complex and other spliceosome assembly factors. Nucleic Acids Res 41, 7485-7500.

Schwer, B., Erdjument-Bromage, H., Shuman, S., 2011. Composition of yeast snRNPs and snoRNPs in the absence of trimethylguanosine caps reveals nuclear cap binding protein as a gained U1 component implicated in the cold-sensitivity of tgs1Delta cells. Nucleic Acids Res 39, 6715-6728.

Schwer, B., Kruchten, J., Shuman, S., 2016. Structure-function analysis and genetic interactions of the SmG, SmE, and SmF subunits of the yeast Sm protein ring. RNA 22, 1320-1328.

Schwer, B., Roth, A.J., Shuman, S., 2017. Will the circle be unbroken: specific mutations in the yeast Sm protein ring expose a requirement for assembly factor Brr1, a homolog of Gemin2. RNA 23, 420-430.

Schwer, B., Shuman, S., 2014. Structure-function analysis of the Yhc1 subunit of yeast U1 snRNP and genetic interactions of Yhc1 with Mud2, Nam8, Mud1, Tgs1, U1 snRNA, SmD3 and Prp28. Nucleic Acids Res 42, 4697-4711.

Schwer, B., Shuman, S., 2015. Structure-function analysis and genetic interactions of the Yhc1, SmD3, SmB, and Snp1 subunits of yeast U1 snRNP and genetic interactions of SmD3 with U2 snRNP subunit Lea1. RNA 21, 1173-1186.

Smith, V., Barrell, B.G., 1991. Cloning of a yeast U1 snRNP 70K protein homologue: functional conservation of an RNA-binding domain between humans and yeast. EMBO J 10, 2627-2634.

Tang, J., Abovich, N., Fleming, M.L., Seraphin, B., Rosbash, M., 1997. Identification and characterization of a yeast homolog of U1 snRNP-specific protein C. EMBO J 16, 4082-4091.

Wilmes, G.M., Bergkessel, M., Bandyopadhyay, S., Shales, M., Braberg, H., Cagney, G., Collins, S.R., Whitworth, G.B., Kress, T.L., Weissman, J.S., Ideker, T., Guthrie, C., Krogan, N.J., 2008. A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. Mol Cell 32, 735-746.

CHAPTER 4

Insights into the contribution of Luc7 subunit from the cryo-EM structure of the *Saccharomyces cerevisiae* U1 snRNP

Structure of Luc7

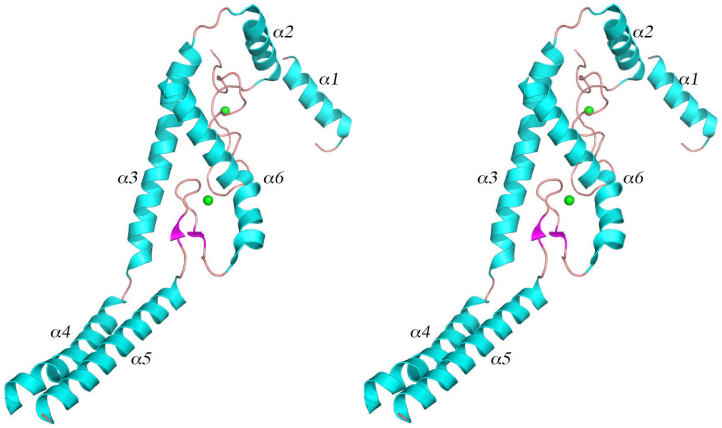
The model of U1 snRNP (Fig. 1.4) was obtained from recently published cryo-EM structures of U1 snRNP, pre-B complex and A-complex of the *S. cerevisiae* spliceosome (Bai et al., 2018; Li et al., 2017; Plaschka et al., 2018). The cryo-EM structures of pre-B complex and A-complex were also associated with the pre-mRNA. These models thus provided some interesting insights to the role of U1 snRNP subunits in spliceosome assembly and U1•5'SS stabilization.

The model of Luc7, a 30kDa protein was obtained from cryo-EM structures of the pre-B complex (pdb 5ZWN) and A-complex (pdb 6G90). The secondary structure elements (obtained from pdb 5ZWN) are aligned over the Luc7 amino acids in Figure 4.1B. Luc7 comprises of six α-helices, and a β hairpin. The tertiary structure of Luc7 is shown in stereo view in Figure 4.1A.

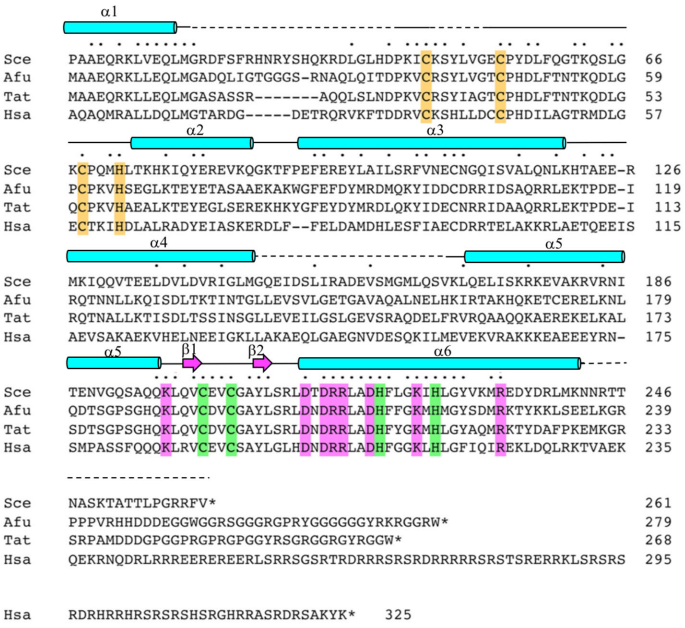
It was previously known that Luc7 is distinguished by two putative zinc finger motifs of CCCH and C2H2 type (Agarwal et al., 2016; Puig et al., 2007). This was confirmed from the model of Luc7 where the ZnF1 ($C^{45}C^{53}C^{68}H^{72}$) motif is seen to contain the zinc binding core in the loop region while the ZnF2 ($C^{201}C^{204}H^{220}H^{226}$) appears to be similar to classical C2H2 zinc finger where two zinc ligands are contributed by conserved residues at the end of a β-strand and the other two zinc ligands are provided by the C-terminus of a α-helix (Krishna et al., 2003) (Fig.4.3 A and B).

Figure 4.1. Structure of Luc7 subunit of *S. cerevisiae* U1 snRNP. (A) A stereo view of the tertiary structure of Luc7 (as observed in pdb 5ZWN) is shown as a cartoon model, with β strands colored magenta and α helices colored cyan. The zinc ions are depicted as a green sphere. (B) Secondary structure elements (colored as in A) are displayed above the *Saccharomyces cerevisiae* (Sce) Luc7 primary structure, which is aligned to the primary structures of *Aspergillus fumigatus* (Afu), *Trichoderma atroviridae* (Tat), and *Homo sapiens* (Hsa) Luc7. Positions of side chain identity/similarity are indicated by (•). Gaps in the alignment are denoted by dashes. The conserved amino acids of the N-terminal involved in zinc-coordination are shaded gold; the amino acids of the conserved C-terminal involved in zinc-coordination are shaded green. Eight other conserved positions at the C-terminus are shaded magenta.

A


α2
α1
α3
α6
α4
α5
α2
α1
α3
α6
α4
α5

B

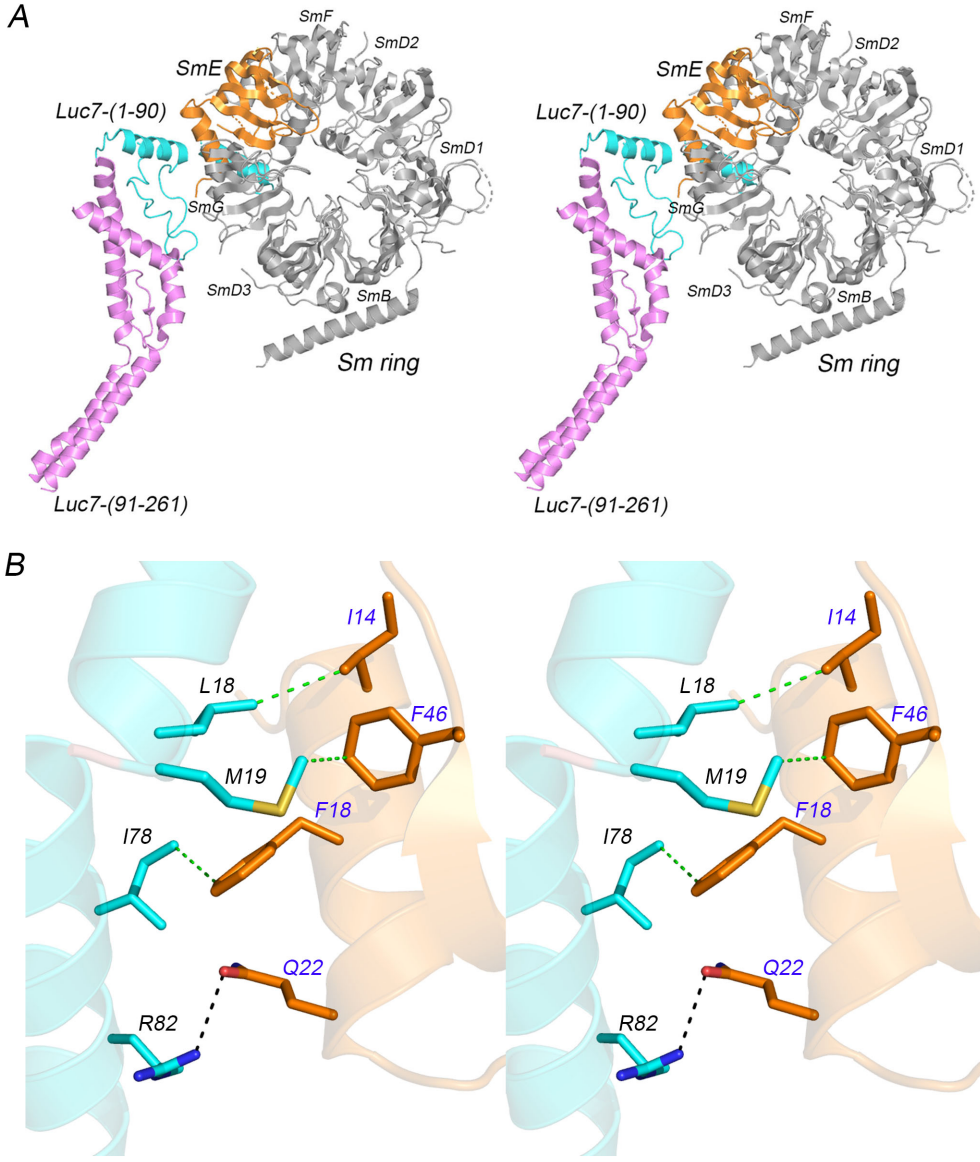

α1
Sce PAAEQRKLVEQLMGRDFSFRHNRYSHQKRDLGLHDPKICKSYLVGECPYDLFQGTKQSLG 66
Afu MAAEQRKLLEQLMGADQLIGTGGGS-RNAQLQITDPKVCRSYLVGTCPHDLFTNTKQDLG 59
Tat MAAEQRKLLEQLMGASASSR-------AQQLSLNDPKVCRSYIAGTCPHDLFTNTKQDLG 53
Hsa AQAQMRALLDQLMGTARDG-----DETRQRVKFTDDRVCKSHLLDCCPHDILAGTRMDLG 57
α2
α3
Sce KCPQMHLTKHKIQYEREVKQGKTFPEFEREYLAILSRFVNECNGQISVALQNLKHTAEE-R 126
Afu PCPKVHSEGLKTEYETASAAEKAKWGFEFDYMRDMQKYIDDCDRRIDSAQRRLEKTPDE-I 119
Tat QCPKVHAEALKTEYEGLSEREKHKYGFEYDYMRDLQKYIDECNRRIDAAQRRLEKTPDE-I 113
Hsa ECTKIHDLALRADYEIASKERDLF--FELDAMDHLESFIAECDRRTELAKKRLAETQEEIS 115
α4
α5
Sce MKIQQVTEELDVLDVRIGLMGQEIDSLIRADEVSMGMLQSVKLQELISKRKEVAKRVRNI 186
Afu RQTNNLLKQISDLTKTINTGLLEVSVLGETGAVAQALNELHKIRTAKHQKETCERELKNL 179
Tat RQTNALLKTISDLTSSINSGLLEVEILGSLGEVSRAQDELFRVRQAAQQKAEREKELKAL 173
Hsa AEVSAKAEKVHELNEEIGKLLAKAEQLGAEGNVDESQKILMEVEKVRAKKKEAEEEYRN- 175
α5
β1
β2
α6
Sce TENVGQSAQQKLQVCEVCGAYLSRLDTDRRLADHFLGKIHLGYVKMREDYDRLMKNNRTT 246
Afu QDTSGPSGHQKLQVCDVCGAYLSRLDNDRRLADHFFGKMHMGYSDMRKTYKKLSEELKGR 239
Tat SDTSGPSGHQKLQVCDVCGAYLSRLDNDRRLADHFYGKMHLGYAQMRKTYDAFPKEMKGR 233
Hsa SMPASSFQQQKLRVCEVCSAYLGLHDNDRRLADHFGGKLHLGFIQIREKLDQLRKTVAEK 235
Sce NASKTATTLPGRRFV* 261
Afu PPPVRHHDDDEGGWGGRSGGGRGPRYGGGGGGYRKRGGRW* 279
Tat SRPAMDDDGPGGPRGPRGPGGYRSGRGGRGYRGGW* 268
Hsa QEKRNQDRLRRREEREREERLSRRSGSRTRDRRRSRSRDRRRRRSRSTSRERRKLSRSRS 295
Hsa RDRHRRHRSRSRSHSRGHRRASRDRSAKYK* 325

Interaction of the Luc7 N-terminus with other U1 snRNP subunits and the effect of N-terminal truncations

I have shown that Luc7 truncations 19-261 and 31-261 were viable at all temperatures but elicited strong synthetic phenotypes in most of the genetic backgrounds (Fig. 2.3). Additionally, they bypassed the requirement for DEAD-box ATPase Prp28 (Fig. 2.4) and showed severe defects in *SUS1* and *YFR045w* pre-mRNA splicing with near absent mature mRNA and increase in the levels of the splicing intermediates or unspliced pre-mRNA (Fig.2.5). The deletion of N-terminal 90 amino acids in Luc7 were lethal.

As observed from the Cryo-EM model (pdb 5ZWN) and shown in Fig. 4.2A, Luc7 N-terminal amino acids 1-90 interact with the SmE subunit of Sm ring. Vander waal's interaction between Luc7 amino acids Leu18, Met19 and Ile78 and SmE amino acids I14, F46 and F18 appear to stabilize the interactions of Luc7 with SmE (Fig. 4.2B). Also, hydrophilic interactions were observed between SmE Gln22 and Luc7 Arg82. Thus, the splicing defects and exacerbated synthetic phenotypes observed for these truncations can be attributed to their role in stabilizing the U1 snRNP complex via its interaction with other U1 snRNP subunits. The bypass of Prp28 requirement in case of N-terminal truncations can be explained by the indirect effect of the N-terminus in stabilizing the U1 snRNA: pre-mRNA duplex.

Figure 4.2. Luc7 contacts SmE subunit of Sm ring. (A) Stereo view of the overall interaction of Luc7 N-terminal amino acids 1-90 (colored cyan) with SmE subunit of Sm ring (colored orange). Other subunits of Sm ring are colored grey. (B) Stereo view of the Luc7 N-terminus highlighting its interactions with SmE subunit of Sm ring. Vander waal's contacts are denoted by green dashed lines and hydrogen bond contacts are denoted by black dashed lines.


A
SmF
SmD2
SmE
Luc7-(1-90)
SmD1
SmG
SmD3
SmB
Sm ring
Luc7-(91-261)
SmF
SmD2
SmE
Luc7-(1-90)
SmD1
SmG
SmD3
SmB
Sm ring
Luc7-(91-261)
B
I14
L18
F46
M19
F18
I78
Q22
R82
I14
L18
F46
M19
F18
I78
Q22
R82

Mutagenesis of the ZnF1 cysteines and histidines

As discussed in Chapter 2, Cys45, Cys53, Cys68 and His72 were mutated to alanine and it was observed that C68A and H72A had no effect on cell growth. However, C45 and C53 were seen to be temperature-sensitive (*ts*) (Fig. 2.1). As seen from the structure these amino acids do not appear to have any contacts with other subunits of U1 snRNP. Moreover, none of the ZnF1 amino acids appear to have interactions with the U1 snRNA or the pre-mRNA. Their functional significance as gleaned from: (i) defective splicing of *SUS1* in *C68A* (Fig. 2.5) and (ii) enhanced synthetic phenotypes in case of *C45A*, *C53A* and *C68A* (Fig. 2.3, Table 4.1) may now be explained due to their imputed role in Zn coordination complex where they play a role in stabilizing a specific tertiary structure that is functionally relevant (Fig. 4.3A). In case of *H72A* where no growth defects or synthetic phenotypes were observed, no other histidine residues in the vicinity appear close enough to be able to replace the coordination of H72 with zinc in order to explain the results. Thus, as discussed previously H72 may be replaced by a water molecule in the metal coordination complex.

Mutagenesis of ZnF2 cysteines and histidines and the role of ZnF2 amino acids in U1•5'SS stabilization

As gauged from previous results ZnF2 motif was particularly interesting such that the Zn binding ligands C201, C204, H220 and H226 were lethal when mutated to alanine (Fig. 2.1, 4.3B). Moreover, eight of the conserved basic and acidic amino acids in this motif though did not result in growth defects (Fig. 2.1) when mutated to alanine exhibited strong functional defects with respect to a reduced splicing efficiency of *SUS1* pre-mRNA (Fig. 2.5), mutational synergies with other splicing components (Fig. 2.3) and a few of them bypassed the Prp28 requirement (Fig. 2.4).

Table 4.1. Summary of synthetic interactions of zinc finger modules of Luc7

Luc7 aa	Zn-coordination	Phenotype of alanine mutant							
			*mud1*Δ	*nam8* Δ	*mud2* Δ	*tgs1*Δ	*cbc2-Y24A*	*prp40-*ΔWW	*Snp1 (1-233)*
C45	ZnF1	ts	lethal	lethal	lethal	lethal	lethal	very sick	lethal
C53		ts	lethal	lethal	lethal	lethal	lethal	very sick	lethal
C68			ts	ts	lethal	lethal	lethal		ts
H72									
C201	ZnF2	lethal							
C204		lethal							
H220		very sick							
H226		lethal							

It was interesting to see that two of the amino acids in this zinc finger motif were involved in stabilizing the U1•5'SS interaction. I observed from the U1 snRNP structure of the pre-B complex (Bai et al., 2018) that Luc7 Asp212 makes a hydrogen bond with G6 ribose O2' in the pre-mRNA, Arg216 makes hydrogen bonds with the U6 base of U1 snRNA and U5 base of pre-mRNA (Fig. 4.4A, Table 4.2). I had previously determined that alanine substitution of R216 can bypass Prp28 requirement with *LUC7- R216A* cells growing as well as wild-type at 37˚ and 34˚C (Fig. 2.4). This observation is now confirmed to be due to the role Arg216 in stabilizing the U1•5'SS interaction. The enhanced synthetic lethality observed in case of *D212A* and *R216A* cells are also now explained in the light of their interactions with U1 snRNA and pre-mRNA thus suggesting their role in spliceosome assembly.

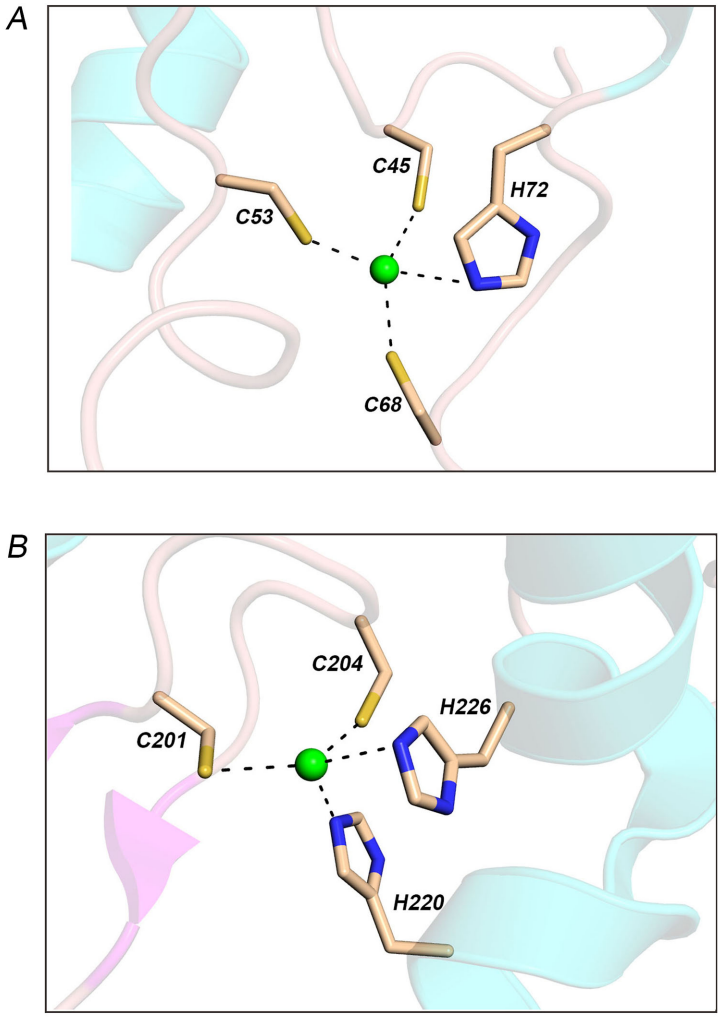


Figure 4.3. Zinc co-ordination by Luc7. (A) and (B) ZnF1 and ZnF2 zinc co-ordinating motifs of Luc7 are represented here as seen from the Cryo-EM structure of Luc7 (pdb 5ZWN). The zinc ion is tetrahedrally coordinated by C45, C53, C68 and H72 in ZnF1 and C201, C204, H220 and H226 in ZnF2. Zinc is represented as a green sphere and co-ordinating cysteines and histidines are represented as stick models with beige carbons.

Figure 4.4. Luc7 ZnF2 interactions with U1 snRNA and pre-mRNA. (A)Stereo view of the modelled Luc7 and its interactions with the segments of U1 snRNA (^{1}AUACUUACCU^{10}) and pre-mRNA (^{1}GGUAUGUAUU^{10}) from the 3.3 Å Cryo-EM structure of the U1 snRNP (PDB ID: 5ZWN). Nucleobases of U1 snRNA are rendered as a stick/cartoon model with gray carbon and gold cartoon trace through the phosphodiester backbone. Nucleobases of pre-mRNA are rendered as a stick/cartoon model with violet carbon and red cartoon trace through the phosphodiester backbone. Luc7 is a cartoon model, colored by domain. (B) Stereo view highlighting the interaction of conserved ZnF2 amino acids (represented as stick model) with U1 snRNP and pre-mRNA. Nucleobases of U1 snRNA and pre-mRNA are colored as in (A). Polar contacts between protein and RNA are denoted by black dashed lines.

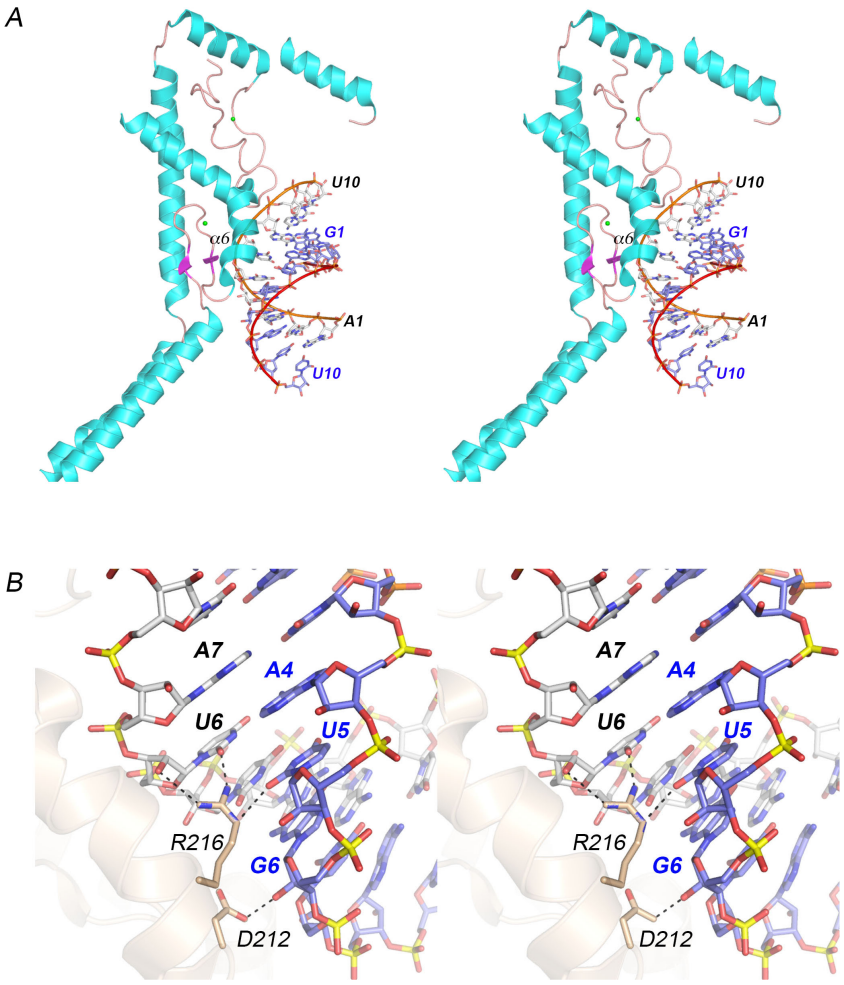

A
U10
α6
G1
A1
U10
U10
α6
G1
A1
U10
B
A7
A4
U6
U5
R216
G6
D212
A7
A4
U6
U5
R216
G6
D212

Table 4.2. Summary of contacts of Luc7 ZnF2 amino acids with U1.5'SS duplex and the effect of their alanine mutations on synthetic genetic interactions

Luc7 aa	Contact		Phenotype of alanine mutant					
	U1 snRNA	pre-mRNA	*mud1*Δ	*nam8*Δ	*mud2*Δ	*tgs1*Δ	*cbc2-Y24A*	*prp40-ΔWW*
D212		D212 Oδ1 with G6 ribose O2'	ts,cs	cs	lethal	cs	cs	
R216	R216 Nε2 with U6 O2	R216 Nε1 with U5 O2	lethal	lethal	lethal	lethal	lethal	very sick

Summary

Insights from cryo-EM model of Luc7 explain the exacerbated synthetic lethal phenotypes and defects in splicing of pre-mRNAs like *SUS1* (which has a nonconsensus 5'SS and branchpoint sequence) caused by otherwise benign mutations in Luc7. The structure highlights the role of the Luc7-N-terminus in stabilizing the U1 snRNP complex via its interaction with other U1 snRNP subunits. Furthermore, the conserved amino acids in the ZnF2 module which showed enhanced synthetic lethal phenotypes and Prp28 bypass were determined from the structure to play a role in stabilizing the U1:5'SS interaction. These observations thus underscore the concept of genetic redundancy and highlighting the essential role of Luc7 in *S. cerevisiae* pre-mRNA splicing.

REFRENCES

Agarwal, R., Schwer, B., Shuman, S., 2016. Structure-function analysis and genetic interactions of the Luc7 subunit of the Saccharomyces cerevisiae U1 snRNP. RNA 22, 1302-1310.

Bai, R., Wan, R., Yan, C., Lei, J., Shi, Y., 2018. Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation. Science 360, 1423-1429.

Krishna, S.S., Majumdar, I., Grishin, N.V., 2003. Structural classification of zinc fingers: survey and summary. Nucleic Acids Res 31, 532-550.

Li, X., Liu, S., Jiang, J., Zhang, L., Espinosa, S., Hill, R.C., Hansen, K.C., Zhou, Z.H., Zhao, R., 2017. CryoEM structure of Saccharomyces cerevisiae U1 snRNP offers insight into alternative splicing. Nat Commun 8, 1035.

Plaschka, C., Lin, P.C., Charenton, C., Nagai, K., 2018. Prespliceosome structure provides insights into spliceosome assembly and regulation. Nature.

Puig, O., Bragado-Nilsson, E., Koski, T., Seraphin, B., 2007. The U1 snRNP-associated factor Luc7p affects 5' splice site selection in yeast and human. Nucleic Acids Res 35, 5874-5885.